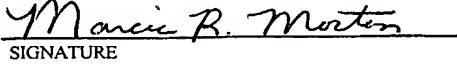


S-31147A

10/088384

**TRANSMITTAL LETTER TO THE UNITED STATES  
DESIGNATED/ELECTED OFFICE (DO/EO/US)  
CONCERNING A FILING UNDER 35 U.S.C. 371**

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/EP00/08994	September 14, 2000	September 16, 1999
TITLE OF INVENTION <i>Transcriptionally Silenced Plant Genes</i>		
APPLICANT(S) FOR DO/EO/US <i>Steimer, Andrea et al.</i>		
Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:		
<p>1. <input checked="" type="checkbox"/> This is a <b>FIRST</b> submission of items concerning a filing under 35 U.S.C. 371.</p> <p>2. <input type="checkbox"/> This is a <b>SECOND</b> or <b>SUBSEQUENT</b> submission of items concerning a filing under 35 U.S.C. 371.</p> <p>3. <input checked="" type="checkbox"/> This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.</p> <p>4. <input type="checkbox"/> The US has been elected by the expiration of 19 months from the priority date (Article 31).</p> <p>5. <input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 371(c)(2))</p> <ul style="list-style-type: none"> <li>a. <input checked="" type="checkbox"/> is attached hereto (required only if not communicated by the International Bureau).</li> <li>b. <input type="checkbox"/> has been communicated by the International Bureau.</li> <li>c. <input type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US).</li> </ul> <p>6. <input type="checkbox"/> An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).</p> <ul style="list-style-type: none"> <li>a. <input type="checkbox"/> is attached hereto.</li> <li>b. <input type="checkbox"/> has been previously submitted under 35 U.S.C. 154(d)(4).</li> </ul> <p>7. <input type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))</p> <ul style="list-style-type: none"> <li>a. <input type="checkbox"/> are attached hereto (required only if not communicated by the International Bureau).</li> <li>b. <input type="checkbox"/> have been communicated by the International Bureau.</li> <li>c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired.</li> <li>d. <input type="checkbox"/> have not been made and will not be made.</li> </ul> <p>8. <input type="checkbox"/> An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).</p> <p>9. <input checked="" type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).</p> <p>10. <input type="checkbox"/> An English language translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).</p>		
<p><b>Items 11 to 20 below concern document(s) or information included:</b></p> <p>11. <input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98.</p> <p>12. <input checked="" type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.</p> <p>13. <input checked="" type="checkbox"/> A <b>FIRST</b> preliminary amendment.</p> <p>14. <input type="checkbox"/> A <b>SECOND</b> or <b>SUBSEQUENT</b> preliminary amendment.</p> <p>15. <input type="checkbox"/> A substitute specification.</p> <p>16. <input type="checkbox"/> A change of power of attorney and/or address letter.</p> <p>17. <input checked="" type="checkbox"/> A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825. <i>+ Paper Copy of Sequence Listings.</i></p> <p>18. <input type="checkbox"/> A second copy of the published international application under 35 U.S.C. 154(d)(4).</p> <p>19. <input type="checkbox"/> A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).</p> <p>20. <input checked="" type="checkbox"/> Other items or information: <i>Certified copy of Priority Document GB9921964.4 Application Data Sheet; International Search Report; PTO-2038</i></p>		

U.S. APPLICATION NO. OR I.D. NO., ETC. (37 CFR 1.4)	INTERNATIONAL APPLICATION NO. PCT/EP00/08994	ATTORNEY'S DOCKET NUMBER S-31147A
21. <input checked="" type="checkbox"/> The following fees are submitted:		CALCULATIONS PTO USE ONLY
<b>BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):</b>		
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO ..... \$1040.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO ..... \$890.00		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO ..... \$740.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) ..... \$710.00		
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) ..... \$100.00		
<b>ENTER APPROPRIATE BASIC FEE AMOUNT =</b>		\$ 890.00
Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e)).		
CLAIMS	NUMBER FILED	NUMBER EXTRA
Total claims	16 - 20 =	x \$18.00
Independent claims	2 - 3 =	x \$84.00
MULTIPLE DEPENDENT CLAIM(S) (if applicable)		+ \$280.00
<b>TOTAL OF ABOVE CALCULATIONS =</b>		\$
<input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.		
<b>SUBTOTAL =</b>		\$
Processing fee of \$130.00 for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(f)).		
<b>TOTAL NATIONAL FEE =</b>		\$
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 per property +		
<b>TOTAL FEES ENCLOSED =</b>		\$ 930.00
		Amount to be refunded:
		charged:
a. <input type="checkbox"/> A check in the amount of \$ _____ to cover the above fees is enclosed.		
b. <input type="checkbox"/> Please charge my Deposit Account No. _____ in the amount of \$ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.		
c. <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. <u>50-1744</u> . A duplicate copy of this sheet is enclosed.		
d. <input checked="" type="checkbox"/> Fees are to be charged to a credit card. <b>WARNING:</b> Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.		
NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.		
SEND ALL CORRESPONDENCE TO:		
 SIGNATURE		
<u>Marcia R. Morton</u> NAME		
<u>46,942</u> REGISTRATION NUMBER		
Customer Number 22847		

10/088584  
S-31147A  
ET327548479US  
JC10 Rec'd PCT/PTO 15 MAR 2002

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re application of:

STEIMER *et al.*

Serial Number: National Stage Application  
of PCT/EP00/08994

Filed: March 15, 2002

For: Transcriptionally Silenced Plant Genes

Art Unit: TBA

Examiner: TBA

Atty Docket: S-31147A

**PRELIMINARY AMENDMENT**

BOX PCT  
U.S. Patent and Trademark Office  
P.O. Box 2327  
Arlington, VA 2202

Sir:

Applicants respectfully request that the above-captioned application be amended as follows in advance of prosecution:

**IN THE SPECIFICATION:**

At page 1, after the title, please insert:

This application is a § 371 of International Application No. PCT /EP00/08994,  
filed September 14, 2000.

**IN THE CLAIMS**

Please amend claim 5 to read as follows:

5. (Amended) The nucleic acid of claim 1 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

S-31147A  
ET327548479US

Please add the following new claims:

14. (New) The nucleic acid of claim 2 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids length in length being at least 85% identical to an aligned component of SEQ ID NO: 10.
15. (New) The nucleic acid of claim 3 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.
16. (New) The nucleic acid of claim 4 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

**REMARKS**

Claim 5 has been amended and new claims 14-16 have been added to eliminate multiple dependency and to ensure that they are in compliance with the rules of U.S. practice. No new matter has been added.

If in the opinion of the Examiner, a telephone conference would expedite the prosecution of the subject application, the Examiner is invited to call the undersigned attorney.

The Commissioner has been authorized to charge the fees associated with the filing of the application that this amendment accompanies and associated papers to Applicant's credit card. If there are any additional fees required, authorization is hereby provided to

37 C.F.R. § 1.17(e)(4) - 37 C.F.R. § 1.17(e)(4)

S-31147A  
ET327548479US

charge such fees to Applicant's Deposit Account No. 50-1744 (in the name of Syngenta Biotechnology, Inc.).

Respectfully submitted,

Syngenta Biotechnology, Inc.  
Patent Department  
P.O. Box 12257  
Research Triangle Park, NC 27709-2257  
(919) 541-8566  
December 21, 2001

*Marcia R. Morton*  
Marcia R. Morton  
Attorney for Applicants  
Reg. No. 46,942

**Version With Markings To Show Changes Made**

The specification has been amended as follows to include the priority data:

--This application is a § 371 of International Application No. PCT /EP00/08994, filed September 14, 2000.--

The claims have been amended as follows:

5. (Amended) The nucleic acid of [any one of claims 1 to 4] claim 1 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

-- 14. (New) The nucleic acid of claim 2 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids length in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

15. (New) The nucleic acid of claim 3 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

16. (New) The nucleic acid of claim 4 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.--

10/088384

Transcriptionally Silenced Plant Genes

The present invention relates to the field of gene expression in plants and in particular concerns gene silencing, a phenomenon frequently observed after integration of transgenes into plant genomes. Comparison of transcriptional gene expression between an *Arabidopsis* line carrying a silent transgene present in multiple copies and its mutant derivative *mom1* impaired in silencing of the transgene revealed two cDNA clones which are expressed in the mutant plants, but not in the parental and not in wild type plants. Both clones are derived from the same family of transcripts which we refer to as TSI (Transcriptionally Silent Information). The disclosed genomic templates encoding TSI are repetitive elements with mainly pericentromeric location and conserved organization among various ecotypes. They are also referred to as TSI. Transcriptional silencing of the genomic TSI templates is specifically released in the mutant. Silencing of said templates is further released in other genotypes known to affect transcriptional gene silencing. Thus, transcription of TSI can be used as a marker to identify a defective silencing pathway in a plant.

Correct balance between activation and silencing of its genetic information is essential for any living cell. A tight control of gene expression is necessary for adaptation to environmental factors, regulation of physiological requirements, and development of differentiated, specialized cell types within a multi-cellular organism. For example differentiation processes involve mitotically heritable changes of gene expression, wherein the acquired states of gene activity gain a certain stability. This stability can be achieved by the strict control of gene activators, by regulation of transcript stability, or by regulating the transcriptional availability of genetic information itself as by stable silencing of selected genetic loci. Silencing has been frequently observed in connection with repression of transgene expression in various experimental systems.

In plants, silencing of transgenic loci limits the reliability of transgenic approaches to improve quality traits. It has been noticed that complex inserts containing rearranged multiple copies of a transgene are particularly prone for gene silencing. Two different mechanisms leading to loss of transgene expression are observed. The first prevents transcription (transcriptional gene silencing or TGS), and the second targets selected transcripts for rapid degradation (posttranscriptional gene silencing or PTGS). Triggers of both processes seem to be similar, since the onset of both types of silencing correlates with redundancy of genetic information, i.e. DNA repeats in case of TGS and RNA

overproduction for PTGS. TGS is meiotically heritable and correlates with DNA template modification manifested by hypermethylation of promoters of silenced genes or with local changes of chromatin structure. In contrast, PTGS is not meiotically transmitted and needs to be reestablished in each sexual generation. PTGS does not require modification of a DNA template, however, increased levels of DNA methylation within the protein-coding region of silenced genes have been observed.

The majority of silencing studies in plant systems deal with silencing of transgenes. There are only a few examples of gene silencing without involvement of transgenic loci. The criteria for TGS susceptibility of genetic information is very poorly understood, and the natural targets of transcriptional silencing in a normal, wild type plant are yet to be discovered. It has been postulated that TGS is a defense system against invasive DNA such as transposable elements but experimental evidence for this hypothesis is lacking.

Within the context of the present invention reference to a gene is to be understood as reference to a DNA coding sequence associated with regulatory sequences, which allow transcription of the coding sequence into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Examples of regulatory sequences are promoter sequences, 5' and 3' untranslated sequences, introns, and termination sequences.

A promoter is understood to be a DNA sequence initiating transcription of an associated DNA sequence, and may also include elements that act as regulators of gene expression such as activators, enhancers, or repressors.

Expression of a gene refers to its transcription into RNA or its transcription and subsequent translation into protein within a living cell.

Any part or piece of a specific nucleotide or amino acid sequence is referred to as a component sequence.

It is the aim of the present invention to provide nucleic acid molecules encoding genetic information which is not expressed, i.e. silenced, in wild type plants but whose expression is turned on in plants which are defective in transcriptional gene silencing. Said molecules can be defined by the formula R<sub>A</sub>-R<sub>B</sub>-R<sub>C</sub>, wherein

- R<sub>A</sub>, R<sub>B</sub> and R<sub>C</sub> indicate component sequences consisting of nucleotide residues independently selected from the group of G, A, T and C or G, A, U and C, wherein  
G is Guanosinmonophosphate,  
A is Adenosinemonophosphate,

T is Thymidinmonophosphate,

U is Uridinmonophosphate and

C is Cytidinmonophosphate;

- $R_A$  and  $R_C$  consist independently of 0 to 6000 nucleotide residues;
- $R_B$  consists of at least 50 nucleotide residues; and
- the component sequence  $R_B$  is at least 80% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

In a preferred embodiment of the present invention  $R_B$  consists of at least 100 nucleotide residues and is at least 85% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

In another preferred embodiment  $R_B$  consists of at least 200 nucleotide residues and is at least 90% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

Specific examples of  $R_B$  are the sequences given in SEQ ID NO: 7, SEQ ID NO: 9 and SEQ ID NO: 27.

Additionally,  $R_A$  or  $R_C$  may comprise one or more component sequences with a length of at least 50 nucleotide residues and at least 90% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

The nucleic acid molecules according to the present invention exist either in the form of DNA or as RNA. Preferred embodiments are genomic DNA, cDNA, plasmid DNA or RNA transcribed therefrom.

Nucleotides 437-2383 of SEQ ID NO: 1 encode a putative open reading frame of 648 amino acids (SEQ ID NO: 10) which in SEQ ID NO: 1 is interrupted by a stop codon spanning nucleotides 1631-1633. Nucleic acids encoding a protein comprising a component sequence of at least 200 amino acids length being at least 85% identical to an aligned component sequence of SEQ ID NO: 10 are a further preferred embodiment of the present invention.

The nucleic acids according to the present invention represent an endogenous target of the transcriptional silencing system. Example 1 describes the cloning of specific embodiments of the present invention from *Arabidopsis*. The preferred size of transcribed RNA is between 1000 and 6000 nucleotides, particularly transcripts of about 1250, 2500, 4700 and 5000 nucleotides, which can be polyadenylated or not. The transcriptionally silent information present in the genome of wild type plants is found to be only expressed in a range of mutants affected in the maintenance of transcriptional silencing. Importantly, not only strains affected in transcriptional silencing through alterations of genome-wide DNA methylation, but also silencing mutants with unchanged methylation levels which do not show striking phenotypic alterations activate TSI, indicating that the release of silencing from endogenous templates does not require loss of methylation.

Initially two independent clones representing RNA which is specifically expressed in silencing mutants have been cloned. Anticipating that in wild type plants there are probably many more DNA templates suppressed by the silencing system, it is remarkable that parts of the two cDNAs cloned are closely related to each other and it is most likely that they are parts of the same transcript. The three main TSI transcripts of 5000, 2500 and 1250 nucleotides all contain a middle element isolated as TSI-A (SEQ ID NO: 5). The 5000 nt and the 2500 nt transcripts additionally enclose the second isolated element TSI-B (SEQ ID NO: 6), which is like TSI-A without protein coding capacity. The 5000 nucleotide long transcript further comprises a 5' extension (SEQ ID NO: 1 which is similar to SEQ ID NO: 2) encoding a putative open reading frame of 648 amino acids (SEQ ID NO: 10). The two 3' extension clones of TSI-A (SEQ ID NO: 3 and SEQ ID NO: 4) contain a region which can be aligned with nucleotides 1-569 of SEQ ID NO: 6 (nucleotides 808-1397 of SEQ ID NO: 3 and nucleotides 819-1365 of SEQ ID NO: 4) closely related to TSI-B (77 % identity). Both the 5000 and the 2500 nucleotide transcripts are polyadenylated, while the most abundant transcript of 1250 nucleotides is absent from the polyA fraction of *mom1* RNA and might be retained in the nucleus.

All RNA species originate from unidirectional transcription, but it is not clear if they represent separate transcriptional units regulated by different promoters or if they are processing products of the same long transcript. A refined analysis of the TSI expression pattern is complicated by the multiplicity of potential chromosomal templates and their location mainly in the pericentromeric areas.

The novel TSI sequences do not reveal any putative function by sequence similarity to protein- or RNA-coding sequences. The only extensive similarity found was to the 3' half of the putative, degenerated retrotransposon *Athila* (Pélissier et al. 1995). The other part of *Athila* directly adjacent to the TSI template region was not reactivated in the silencing mutant. This suggests that epigenetic transcriptional silencing in *Arabidopsis* is not directed towards retrotransposons in general, although its targets may have originated from transposition events. This is further supported by the lack of transcriptional reactivation of other *Arabidopsis* retroelements, e.g. the *Ta* superfamily (Konieczny et al. 1991). Therefore, only specific pericentromeric repeats seem to be under epigenetic control, in the same way that only a subset of transgenic loci is susceptible to silencing. The existence of remnants of transposons is probably due to their chromosomal location rather than to sequence specificity, since degenerated retroelements have repeatedly been found in centromeric locations in fungi and plants.

One of the features proposed as a prerequisite for centromere function is late replication of the heterochromatic centromeres and pericentromeric areas in *Schizosaccharomyces pombe* and higher eukaryotes. If this was also true for *Arabidopsis* centromeres, undue loosening of suppressive chromatin leading to TSI expression could cause disturbances in mitosis, which would result in severe phenotypes. However, the *mom* mutant plants exhibit no abnormalities suggesting mitotic disorders. Therefore, transcriptional reactivation of some usually silent pericentromeric repeats, such as described here, does not impair their putative function. Alternatively, their silencing may be important under a specific, still undefined condition or on a longer time scale.

Finally, TSI expression is observed in cells growing for a long time in suspension culture. No release of TSI silencing is observed in any tissue of *Arabidopsis* wild type plants, including freshly initiated callus cultures. This suggests that an escape from the silencing control is not correlated primarily with dedifferentiation but could be the result of prolonged selection for fast growing dedifferentiated cells. Such a loss of silencing control could also underlie the accumulation of somaclonal variation during prolonged culture and resembles the situation in cells of actively proliferating carcinomas.

Nucleic acids according to the present invention are particularly useful in selecting plants which compared to wild type *Arabidopsis* plants of all available *Arabidopsis* ecotypes are

impaired in transcriptional gene silencing. A method allowing to select such plants comprises

- a) separately preparing RNA of a series of plants;
- b) probing said RNA preparations with a nucleic acid according to the present invention; and
- c) identifying a plant whose RNA hybridizes with said nucleic acid.

In a preferred embodiment the probing step is performed after size fractionation of the RNA preparation by gel electrophoresis. For detection the probe is either radioactively labeled or labeled by other chemical modifications.

In another preferred embodiment of said method the step of probing consists of hybridizing the RNA with an oligonucleotide primer, extending said primer by reverse transcription and subsequent PCR amplification of the DNA generated using oligonucleotide primers specific for SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 27. Plants which allow for the amplification of DNA fragments flanked by the oligonucleotide primers are identified as plants whose RNA hybridizes to the nucleic acid according to the invention.

Having available nucleotide sequence information of a genomic region, which is not expressed, i.e. transcriptionally silenced, in a wild type plant, allows to produce DNA representing at least part of a gene necessary to maintain silencing of this genomic region. Preferably the complete gene is produced. A corresponding method of production comprises

- (a) mutagenizing wild type cells or plants by randomly inserting into their genomes a DNA tag with known sequence;
- (b) identifying mutants of said cells or plants which express RNA that is not expressed in wild type cells or plants;
- (c) cloning genomic DNA surrounding or close to the insertion site of the DNA tag;
- (d) screening a genomic library of wild type cells or plants with the piece of genomic DNA obtained in process step (c) or a part thereof;
- (e) identifying clones comprising at least part of the gene affected by the insertion of the DNA tag; and
- (f) further processing the clones obtained in step (e) using recombinant DNA techniques.

In plant cells and plants mutagenesis is preferably achieved performing T-DNA insertion mutagenesis (Dilkes and Feldmann, 1998), or transposon tagging using the En/I (Pereira and Aarts, 1998) or the Ac/Ds system (Long and Coupland, 1998) as described in *Arabidopsis* protocols edited by Martinez-Zapater and Salinas, 1998. Other known physical or chemical methods of mutagenesis such as fast neutron irradiation or EMS mutagenesis (Feldmann et al., 1994) might require adaptation of the above method, but can be used for the production of equivalent DNA involved in the maintenance of silencing as well. A convenient way to identify RNA that is expressed in mutant cells or plants but not in wild type cells or plants is reverse transcription of said RNA and subsequent PCR amplification of the generated DNA using oligonucleotide primers specific for said DNA (RT-PCR). This allows to pool the RNA of upto 1000 mutants which increases the speed of the identification step considerably.

The methods described above can be further elaborated and developed into a kit for the identification of plants impaired in transcriptional gene silencing. Such a kit necessarily comprises

- a) a nucleic acid according to the present invention conveniently labeled to be used as a hybridization probe or
- b) an oligonucleotide primer for reverse transcription of RNA and an oligonucleotide primer specific for a nucleic acid according to the present invention.

The oligonucleotide primer for reverse transcription can be a poly T primer or an oligonucleotide primer specific for a nucleic acid according to the present invention. The primers specific for nucleic acids according to the present invention are designed to allow PCR amplification of DNA templates characterized by the nucleotide sequences disclosed in the present invention.

### **EXAMPLES**

#### **Example 1: Differential mRNA screening and cloning of *Arabidopsis* TSI sequences**

Total RNA of the mutant line *mom1* (Amedeo et al, 2000) and its parental line A is isolated according to Goodall et al. (1990) using 2 g fresh weight of two-week-old seedlings.

Polyadenylated RNA is obtained using Dynabeads Oligo (dT)<sub>25</sub> (Dynal). 2 µg of poly(A) RNA is used for suppression subtractive hybridization (SSH, Diatchenko et al, 1996) using the PCR-Select cDNA subtraction kit (Clontech) according to the suppliers' instructions. cDNA

derived from the mutant line *mom1* is used as tester and cDNA derived from the parental line A as driver cDNA population. The subtracted library is cloned into vector pCR2.1 (Invitrogen). 500 individual bacterial cultures from this library are grown according to the manual of the PCR-Select differential screening kit (Clontech). To reduce the number of false positive clones, the library is primarily screened by inverted Northern blots as described by von Stein et al. (1997). Twelve among the 500 primarily selected cDNA clones show increased abundance upon hybridization with labeled *mom1* cDNA. Direct Northern blot analysis comparing total RNA of the wild type and the mutant line with these 12 cDNAs as probes reveal a striking genotype-dependent differential expression for two of them. Said clones are sequenced using conventional rhodamine or dRhodamine dye terminators from PE Applied Biosystems and a Perkin-Elmer GeneAmp PCR system 2400, 9600 or 9700 thermocycler. The sequence reactions are analyzed using an ABI PRISM 377 DNA sequencer. The cDNA clones are named TSI-A (903 bp, SEQ ID NO: 5) and TSI-B (614 bp, SEQ ID NO: 6). Both are abundant in the *mom1* RNA but are undetectable in Arabidopsis line A and wild type Arabidopsis. No consistent differential expression between mutant and wild type is observed for the remaining 10 cDNAs.

5' and 3' extension reactions are performed using Clontech's Marathon Kit according to the manufacturer's instructions. Sequence specific primers are

TA-F1: 5'-TGGTTCACCAAGATAAGCTCAGTGCCCTC-3' (SEQ ID NO: 11) and

TA-F2: 5'-CTTCAGACTGGATAGGACTAGGTGGCG-3' (SEQ ID NO: 12, nested primer), for the 3'extension reaction and

TA-R1: 5'-CGCCCACCTAGTCCTATCCAGTCTGAAG-3' (SEQ ID NO: 13) and

TA-R2: 5'-CGCATCAAACAACTAACAAACGAGGGCAC-3' (SEQ ID NO: 14, nested primer). for the 5'extension.

PCR amplification products are cloned into vector pCR2.1 (Invitrogen). Individual bacterial cultures are grown and subjected to colony PCR as described in the manual of Clontech's PCR-Select Differential Screening Kit, with the primer combinations used to create the extension reactions (Marathon Adapter primer Ap1 (Clontech) combined with TA-R2 or TA-F2 for the 5'- or 3' extension reactions, respectively). To screen for positive TSI-A extension clones, the PCR products are blotted and hybridized to TSI-A. All PCR reactions for cloning procedures are performed with a polymerase mix performing proofreading activity (Advantage cDNA PCR Kit, Clontech).

Since only two transcripts are detected in the polyadenylated RNA fraction of *mom1* plants, this RNA is used for 5' and 3' extensions reactions starting from the TSI-A sequence. Two clones each are analyzed at the nucleotide sequence level. The 5' extension yields inserts of 2512 bp (clone a, SEQ ID NO: 1) and 1997 bp (clone b, SEQ ID NO: 2), which after alignment are 97 % identical to each other. The clones from the 3' extension have a length of 1682 bp (clone c, SEQ ID NO: 3) and 1652 bp (clone d, SEQ ID NO: 4) and are 94 % identical. Interestingly, both 3' extension clones of TSI-A contain a region of 569 bp closely related to TSI-B (77 % identity). This explains the detection of similar RNA species on Northern blots with TSI-A and TSI-B as probes and their hybridization to the same YAC and BAC clones, and suggests that TSI-A and TSI-B are part of the same polyadenylated transcript species expressed in the *mom1* mutant. To confirm that the 5' extensions of TSI-A are indeed part of the TSI transcripts, a *mom1* Northern blot is probed with a cDNA fragment close to the 5' end of the extension (probe ORF corresponding to nucleotides 943-1334 of SEQ ID NO: 1). Interestingly, only the about 5000 nt long transcripts in the poly(A) fraction hybridize to this probe. Since this class of transcripts hybridizes to TSI-A and TSI-B, the 5000 nt transcripts are probably produced from templates containing a particular order of all three sequence elements.

#### Example 2: Northern and Southern blot analysis and library screens

Total RNA is either isolated as described by Goodall et al. (1990) or by the RNeasy Plant Mini Kit (Qiagen) according to the suppliers' instructions. For Northern blot analysis, the RNA is electrophoretically separated after denaturation by glyoxal in a 1.5 % agarose gel in phosphate buffer (pH 7) and blotted to nylon membranes (Hybond N, Amersham) using standard protocols. The Boehringer molecular weight marker I is used as a size standard. For Southern blot analysis, genomic DNA is isolated according to Dellaporta et al. (1983) and separated electrophoretically after endonucleolytic digestion. DNA fragments are transferred to nylon membranes (Hybond N, Amersham) according to standard procedures. Hybridization and washing of Northern and Southern blots and the filters with the YAC library is performed according to Church and Gilbert (1984). Probes are labeled with [ $\alpha$ -<sup>32</sup>P]-dATP by random prime DNA polymerization (Feinberg and Vogelstein, 1983) and exposed to X-ray sensitive film (Kodak X-OMAT AR).

Hybridization of Northern blots using total RNA prepared from 2-week-old seedlings with TSI-A of *mom1* visualizes four major transcripts with sizes of approximately 5000, 4700,

2500 and 1250 nucleotides. A TSI-B probe detects mainly two transcripts of 5000 nt and 2500 nt. Interestingly, the polyadenylated fraction of *mom1* RNA contains only the transcripts of 5000 nt and 2500 nt hybridizing to both cDNA probes (TSI-A and TSI-B). From the sizes of TSI transcripts detected it is obvious that both TSI clones represent only partial cDNAs. TSI expression is meiotically heritable and persists through 6 selfed generations of *mom1* with the same pattern of transcripts.

To examine TSI expression in other genotypes known to affect gene silencing, total RNA of several *Arabidopsis* mutant and transgenic strains known to be affected in gene silencing is probed with TSI-A. All the *som* mutants *som1* to *som8*, described by Mittelsten Scheid et al (1998) to be impaired in the maintenance of transcriptional silencing similar to *mom1*, show a high level of TSI-A expression. The mutation *ddm1*, originally identified to have decreased DNA methylation (Vongs et al. 1993), and later revealed to release transcriptional gene silencing from different loci (Mittelsten Scheid et al, 1998; Jeddeloh et al., 1998) also shows a high level of TSI-A expression. TSI-A expression is also expressed in a transgenic line described by Finnegan et al (1996) which shows decreased DNA methylation due to overexpression of DNA methyltransferase antisense mRNA as well as in a further *Arabidopsis* mutant affected in the DNA methyltransferase gene (said mutant is referred to as *ddm2* and has been provided by Eric Richards). Moreover, the silencing mutants *hog1* and *sil1*, but not *sil2* described by Furner et al (1998) express sequences hybridizing to TSI-A. Importantly, mutations affecting posttranscriptional gene silencing such as *sgs1* and *sgs2* described by Elmayan et al (1998) and *egs1* described by Dehio and Schell (1994) do not express RNA which hybridizes to TSI-A.

Comparison of patterns of TSI-A expression in the different genotypes reveals genotype specific differences in the stoichiometry of the different RNA species. *mom1* plants reveal a different expression pattern of TSI-A and TSI-B as compared to *som1* plants. These results indicate that a particular genetic deficiency in the transcriptional silencing system leads to a differential but specific activation of TSI templates. However, we observe variation in these activation patterns between different sources of plant material and different RNA preparations. Therefore, it is possible that patterns of TSI expression are more flexible and probably also controlled by still unknown factors acting in the mutant background, or by different stabilities among the transcript populations.

TSI-A and TSI-B are used as probes for Southern blots to determine the source of TSI transcripts and the organization of their template(s). The blots reveal that multiple copies of

TSI-A- and TSI-B-homologous sequences are present in the genome of *Arabidopsis*. Copy numbers are assessed by reconstruction experiments to approximately 130-300 copies of TSI-A.

To examine the degree of evolutionary conservation of the TSI arrangement, DNA of five *Arabidopsis* ecotypes (Zürich, Columbia, Landsberg erecta, Wassilevskija, C24) is compared by Southern blot analysis and hybridization to the TSI-A probe. Genomic DNA is digested with *Dra*I that has a single recognition site within TSI-A and *Ssp*I that does not cut within TSI-A. A significant conservation of the TSI-A pattern among different ecotypes is observed, with two main *Dra*I repeats of 4 kb and 1.3 kb and two major *Ssp*I fragments of 11 kb and 4 kb. Some minor differences specific for a particular ecotype indicate a limited genetic polymorphism within TSI-A. Probing the same membrane with TSI-B reveals complex banding patterns different in each ecotype which might indicate a lower degree of conservation for TSI-B, although the differences of the Southern blot patterns between TSI-A and TSI-B can also be explained if TSI-A is an internal part of a longer repeated element, and TSI-B is located proximal to a flank between repeated elements and variable single copy DNA regions.

After hybridizing TSI-A and TSI-B to the CIC YAC library covering 4 genome equivalents and 92% of the *Arabidopsis* genome sixty-two CIC clones out of 1152 turn out to hybridize with the TSI-A probe. Twenty-six of these contain also the pericentromeric 180-bp-repeat, 7 contain 5S RNA genes known to be located in the vicinity of a centromere, and 16 clones contain other markers that map close to centromeres. Only 4 of these clones map outside of centromeric regions. Similar mapping of TSI-B results in hybridization to all TSI-A positive CIC clones, with additional 7 clones hybridizing to TSI-B only. Thus both TSI repeats are concentrated in the pericentromeric regions of *Arabidopsis* chromosomes.

After hybridizing TSI-A and TSI-B to a cDNA library prepared with RNA isolated from *mom1* mutant plants, 22 hybridizing clones are further analyzed by sequence analysis. RNA is isolated from 2-week-old seedlings of the *mom1* mutant plant according to Goodall et al. (1990). The cDNA library is prepared using the Uni-ZAP XR library construction kit (Stratagene) according to the manufacturer's protocol. cDNA fragments larger than 500 bp are selected using the cDNA size fractionation columns from Gibco. 7 clones contain SEQ ID NO: 7 (TSI-A-15), 5 clones contain SEQ ID NO: 8 (TSI-A-2) part of which is identical to SEQ ID NO: 4.

**Example 3: Database searches**

Sequence analysis is performed using the GCG software (Wisconsin Package Version10.0, Genetics Computer Group (GCG), Madison, Wisconsin). For identity searches, GenEMBL,

the *Arabidopsis thaliana* database (<http://genome-www3.stanford.edu/>), Kazusa

*Arabidopsis* opening site (KAOS; <http://zearth.kazusa.or.jp/arabi/>) and Swissprot are used.

Peptide sequences are analysed by GeneQuiz (<http://columba.ebi.ac.uk:8765/gqsrv/submit>) or Expasy (<http://www.expasy.ch/>).

Searches within the genomic sequence databases GenEMBL, *Arabidopsis thaliana*

Database, and KAOS confirm the presence of multicopy sequences related to TSI-A and

TSI-B, which are distributed over all five chromosomes of *Arabidopsis thaliana*. Importantly, very often single BAC clones contain sequences homologous to both cDNA clones. In some cases, TSI-A and TSI-B related sequences are found more than once on the same BAC

clone, suggesting that TSI-A and TSI-B belong to a clustered repetitive element.

The significant sequence heterogeneity between the cDNA classes and duplicates of the 5' and 3' extensions of TSI-A indicate that they originate from different activated repeats. To facilitate the data base search for a possible genomic template of the 5000 nt transcript among the multiple related copies, the overlapping cDNA sequences are combined to form a continuous 4860 bp sequence of "virtual" cDNA (SEQ ID NO: 9) which is used to search the *Arabidopsis* genomic sequence databases. A particular BAC clone (TAMU BAC T6C20, accession number AC005898) has 91 % identity to the combined cDNA sequence. Further, the search uncovers a chromosomal DNA stretch (BAC F7N22, accession number AF058825) 99% identical to the abundant cDNA A-15 of Example 2 (SEQ ID NO: 7). The genomic sequence of the transcribed region 5' to the region defined by SEQ ID NO: 7 is given in SEQ ID NO: 27. It is identical to nucleotides 65081-68202 of BAC F7N22. Both sequences are located at the pericentromeric region of chromosome five. The TSI sequence defined by nucleotides 65080 to 70370 on BAC F7N22 is 54 % identical to the retrotransposon-like repeat named *Athila*. The identity of this sequence as a retrotransposon is deduced from *Arabidopsis* genome sequences around heterochromatic regions that are marked by the presence of 180 bp satellite repeats. The 10.5 kb sequence of *Athila* has several characteristics of a retroelement, like long terminal repeats (LTR), a polypurine track (PPT) and a primer binding site (PBS) for tRNA priming of the reverse transcriptase, but its open reading frames do not share homology with proteins known to be

involved in transposition.

The TSIs map to the 3' terminal part of *Athila*. TSI-A covers a part of the 3' non-coding region of the putative retrotransposon and TSI-B corresponds to the PPT and a part of the 3' LTR. The sequence of the 5' TSI-A extension encodes a possible open reading frame of 648 amino acids length (SEQ ID NO: 10) with 51 % identity to 604 amino acids of the ORF2 deduced for *Athila*. The sequence coding for this ORF is also present on the TAMU BAC T6C20, however, the ORFs encoded by the two cDNAs clone a (SEQ ID NO: 1) and clone b (SEQ ID NO: 2) and the BAC are interrupted by translational stop codons after 398 amino acids (clone a), 83 amino acids (clone b) and 46/465/496/499/549 amino acids respectively (BAC T6C20). The ORF2 sequence present on BAC F7N22 is highly degenerated by five deletions of 2-31 bp and five insertions of 3-10 bp. This further supports the assumption that this sequence is derived from a putative but degenerated retrotransposon. Data base searches for proteins similar to the potential product of the 648 amino acids ORF do not yield significantly similar polypeptides neither to proteins usually encoded by retroelements nor to any other known polypeptides.

#### Example 4: RNase protection assays

RNase protection assays are performed according to Goodall et al. (1990) with minor modifications. To assay the direction of TSI transcription, the pCR2.1 based plasmid containing the TSI-A insert is cut by EcoRI creating a fragment of 781 bp which is ligated into the vector pGEM-7Zf(+) (Promega). To map the 5' transcription start, the probe is generated by amplifying the BAC F7N22 region between positions 64929 and 65567 and inserting the product into the pGEM-7Zf(+) vector (Promega). Labeled probes are synthesized by *in vitro* transcription of the linearized plasmid in the presence of [ $\alpha$ -<sup>32</sup>P]-UTP using T7 polymerase (Promega) or Sp6 polymerase (Boehringer) and purified by electrophoresis (Goodall et al., 1990). Single stranded RNA is cleaved by either 4  $\mu$ g RNase A and 0.6 U RNase T<sub>1</sub> (RNase A/T assay) or by 20 U RNase T<sub>1</sub> (RNase T assay). Protected fragments are separated on a denaturing 6 % polyacrylamide gel. The dried gel is exposed to a PhosphorImager screen (Molecular Dynamics) and to X-ray sensitive film.

To determine the polarity of TSI transcription, RNase T and RNase A/T protection assays are performed with TSI-A probes of opposite polarity. TSI-A sequences are used as probes since TSI-A is present in all transcripts detected on Northern blots. There is no evidence for protection of the probe corresponding to the sense strand. This suggests the lack of any

TSI antisense RNA and a unidirectional transcription of the TSI templates. Interestingly, RNase digestions with a TSI-A antisense probe creates a complex pattern of TSI-A protected bands. This suggests that many different but related RNAs hybridize to the probe. Since a fragment of 781 nt as expected for the protection of the entire TSI-A probe is visible, it can be concluded that TSI-A is part of an activated transcript throughout and not an artifact generated by template switch during the SSH procedure. Furthermore, some of the protected TSI-A fragments are clearly more abundant than others, suggesting either a structural conservation of particular regions of TSI-A within related RNAs, or alternatively a higher abundance of certain transcript subspecies.

The sequence information of BAC F7N22 is used to determine the position of the transcription start for the longest TSI transcript. An antisense RNA probe for RNase A/T protection is produced spanning the 638 nucleotides between positions 64929 and 65567. The probe is hybridized with total RNA from *ddm1*, *som7* and *mom1*. In all RNA preparations, a fragment of approximately 480 nt ( $\pm 10$  nt) is protected and allows positioning of the TSI transcription start on BAC F7N22 to 65087 ( $\pm 10$  nt) in different mutants.

#### Example 5: Reverse Transcription PCR (RT PCR)

Reverse transcription is performed with 1  $\mu$ g total RNA from *mom1* in the presence of 1 mM dNTPs, 4-20 U RNasin (Promega), 1x AM RTase buffer (Boehringer) and 25 U AM reverse transcriptase (Boehringer) at 37°C for 1 hour, followed by heat inactivation of the reaction mixture. As template for PCR 50 ng reverse transcribed RNA primed by gene specific antisense primers (BA-R1, BA-R2, AT-R1, and TA-R1, see below) or 100 ng genomic DNA or 100 ng cDNA are used. PCR is started with 3 min denaturation at 94°C, followed by 30 amplification cycles (denaturation at 94°C/30 sec, annealing at 62°C/30 sec, and elongation at 72°C/30 sec) in the presence of 0.2 mM dNTPs, 0.4  $\mu$ M forward and reverse primers, 1x Taq DNA polymerase buffer (Boehringer) and 0.25 U Taq DNA polymerase (Boehringer). The nucleotide sequences of the primers used for RT-PCR are:

AT-F1: 5' -CGATAACATCGACCGTATTGCTCGCC-3' (SEQ ID NO: 15)

AT-R1: 5' -AACTAGCTCCCATCCGTCTTCGACATCC-3' (SEQ ID NO: 16)

AT-F2: 5' -TGCATCACACCGGATTGGATTGAC-3' (SEQ ID NO: 17)

AT-R2: 5' -TGTTCCCTGAACCATAAGCAATGAGAACC-3' (SEQ ID NO: 18)  
BA-F1: 5' -CAAACAGACAGAGTGTGGCCCACCACC-3' (SEQ ID NO: 19)  
BA-R1: 5' -AAGAGAGGGAGAAGGCAGTGGCGTGAG-3' (SEQ ID NO: 20)  
BA-F2: 5' -TGCAAACCCACAGGACCAAGTCTACCC-3' (SEQ ID NO: 21)  
BA-R2: 5' -ACAGATGGTGATAGCGTGAGCGGTGGC-3' (SEQ ID NO: 22)  
F7-F1: 5' -TCAACCTTTGCCCAACAACCAACTC-3' (SEQ ID NO: 23)  
F7-R1: 5' -TCTCCATCCACGCTTCCTGAATGTCC-3' (SEQ ID NO: 24)  
GS-F1: 5' -GGAGAAGGAAGCTGAAATCATATTGTGG-3' (SEQ ID NO: 25)  
GS-R1: 5' -ATGATGATCCTAAGTCTACCCCTTGCAC-3' (SEQ ID NO: 26)

As a positive control for the PCR reactions, the TA-F1 and TA-R1 primers are used. The reverse transcriptase region of the pol gene of *Arabidopsis* Ty1/copia-like retrotransposon family is amplified as described by Konieczny (1991) with immaterial modifications.

Since the nucleotide sequence of the TSI transcripts TSI-A and TSI-B is related to the nucleotide sequence of the 3' half of retrotransposon-like element *Athila* including the second ORF and the 3' LTR, we examined by RT-PCR whether transcription of the 5' part of *Athila* including the first ORF is activated in *mom1* plants. Five primer pairs are chosen (BA-F1 - BA-R1; BA-F2 - BA-R1; AT-F1 - AT-R1; AT-F2 - AT-R2, F7-F1 - F7R1) according to the sequence information about *Athila* and the related parts of BAC T6C20 and BAC F7N22. All primer combinations amplified the expected products from genomic template DNA, but no PCR product could be obtained from *mom1* RNA, regardless, whether cDNA synthesis was started from an *Athila*- or BAC-specific reverse primer or from polyT-primed cDNA (data not shown). Activation of TSI therefore is limited to sequences related to the 3' part of *Athila*.

The two classes of isolated cDNAs share only approximately 50% identity with *Athila*. To directly address the question of whether *Athila* is expressed, RT-PCR experiments are performed with *Athila*-specific primers (GS-F1, GS-R1) in the TSI homologous region. However, the corresponding fragment cannot be amplified from RNA of *mom1* seedlings, suggesting that only a subset of *Athila*-like sequences but not the *Athila* element itself is reactivated in the mutant background.

To investigate whether other retroelements are transcriptionally activated in *mom1*, degenerated primers in a conserved region of the reverse transcriptase gene used to clone

and describe the Ta superfamily of *Arabidopsis* retrotransposons (Konieczny et al., 1991) are used to investigate, if members of the Ta family are transcribed in the mutant background. Although the expected 268 bp fragment can be amplified from genomic DNA, no amplification is achieved in RT-PCR with *mom1* RNA as template. This indicates, that, in spite of the TSI homology to retrotransposons, these elements are not generally activated in the *mom1* mutant.

**Example 6: TSI expression after application of stress (salinity, UV-C, pathogen)**

Induction of TSI upon UV-C is tested on Northern blots with RNA samples from 1 week-old seedlings subjected to UV-C treatment of 1 kJ/m<sup>2</sup> or 5 kJ/m<sup>2</sup> which are collected at several time points within 1 hour (Revenkova et al., 1999). The effect of osmotic stress is tested on Northern blots with RNA from one-week-old seedlings that are transferred for 24 hours to medium with NaCl concentrations of 0, 0.04, 0.08 and 0.12 M (Albinsky et al., 1998). To test TSI expression upon pathogen stress, RNA of 3-week-old seedlings either mock treated or infected with *Peronospora* is analysed by Northern blot analysis. To verify the appropriate pathogen response, induction of PR1 expression is monitored by reprobing the membrane with a PR1 probe.

In young seedlings (2 weeks old) and in different tissues of mature wild type plants (roots, shoots, leaves, flowers, siliques), TSI expression cannot be detected. The application of various stress treatments namely elevated salinity, UV-C, or pathogen infection, does not activate TSI in wild type plants. TSI expression is also not detected in freshly initiated callus cultures, and transcriptional suppression of TSI is stable even after several *in vitro* passages of the callus culture. However, the only exception so far are cells derived from wild type *Arabidopsis* (literature) growing for a long time in suspension culture. These cells express TSI-A, indicating release of TSI silencing under these conditions.

**REFERENCES**

- Albinsky et al, *Plant J.* **17:** 73-82, 1999
- Amedeo et al, *Nature* **405:** 203-206, 2000
- Church and Gilbert, *Proc Natl Acad Sci USA* **81:** 1991-1995, 1984
- Dehio and Schell, *Proc Natl Acad Sci USA* **91:** 5538-5542, 1994
- Dellaporta et al, *Plant Mol Biol Rep.* **1:** 19-21, 1983
- Diatchenko et al, *Proc Natl Acad Sci USA* **93:** 6025-6030, 1996
- Dilkes and Feldmann, *in* Arabidopsis protocols, Methods in Molecular Biology **82** (Ed. Martinez-Zapater and Salinas), Humana Press, Totowa, New Jersey, pp 339-351, 1998
- Elmayan et al, *Plant Cell* **10:** 1747- 1757, 1998
- Feinberg et al, *Anal Biochem.* **132:** 6-13, 1983
- Feldmann et al, *in* Arabidopsis (Ed. Sommerville and Meyerowitz), CSHL Press, New York, pp 137-172, 1994
- Finnegan et al, *Proc Natl Acad Sci USA* **93:** 8449-8454, 1996
- Furner et al, *Genetics* **149:** 651-662, 1998
- Goodall et al, *Methods Enzymol.* **181:** 148-161, 1990
- Jeddeloh et al, *Genes Dev.* **12:** 1714-1725, 1998
- Konieczny et al, *Genetics* **127:** 801-809, 1991
- Long and Coupland *in* Arabidopsis protocols, Methods in Molecular Biology **82** (Ed. Martinez-Zapater and Salinas), Humana Press, Totowa, New Jersey, pp 315-328, 1998
- Mittelsten Scheid et al, *Proc Natl Acad Sci USA.* **95:** 632-637, 1998
- Pellisier et al, *Plant Mol Biol.* **29:** 441-452, 1995
- Pereira and Aarts, *in* Arabidopsis protocols, Methods in Molecular Biology **82** (Ed. Martinez-Zapater and Salinas), Humana Press, Totowa, New Jersey, pp 329-338, 1998
- Revenkova et al, *EMBO J.* **18:** 490-499, 1999
- von Stein et al, *Nucleic Acids Res.* **25:** 2598-2602, 1997
- Vongs et al, *Science.* **260:** 1926-1928, 1993

**What is claimed is:**

1. A nucleic acid having the formula R<sub>A</sub>-R<sub>B</sub>-R<sub>C</sub>, wherein
  - R<sub>A</sub>, R<sub>B</sub> and R<sub>C</sub> constitute component sequences consisting of nucleotide residues independently selected from the group of G, A, T and C or G, A, U and C, wherein G is Guanosinmonophosphate,  
A is Adenosinemonophosphate,  
T is Thymidinmonophosphate,  
U is Uridinmonophosphate and  
C is Cytidinmonophosphate;
  - R<sub>A</sub> and R<sub>C</sub> consist independently of 0 to 6000 nucleotide residues;
  - R<sub>B</sub> consists of at least 50 nucleotide residues; and
  - the component sequence R<sub>B</sub> is at least 80% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.
2. The nucleic acid of claim 1, wherein R<sub>B</sub> consists of at least 100 nucleotide residues and is at least 85% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.
3. The nucleic acid of claim 1, wherein the nucleic acid sequence of R<sub>B</sub> is the sequence given in SEQ ID NO: 7, SEQ ID NO: 9 or SEQ ID NO: 27.
4. The nucleic acid of claim 1, wherein R<sub>A</sub> or R<sub>C</sub> comprise one or more additional component sequences with a length of at least 50 nucleotide residues and at least 90% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.
5. The nucleic acid of any one of claims 1 to 4 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids length being at least 85% identical to an aligned component sequence of SEQ ID NO: 10.

6. A method of selecting a plant which compared to a wild type plant is impaired in transcriptional gene silencing, comprising
  - a) separately preparing RNA of a series of plants;
  - b) probing said RNA preparations with a nucleic acid according to claim 1; and
  - c) identifying a plant whose RNA hybridizes with said nucleic acid.
7. The method of claim 6, wherein process steps b) and c) comprise reverse transcription of the RNA and subsequent amplification of the generated DNA using oligonucleotide primers specific for SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 or SEQ ID NO: 27.
8. A method of producing DNA representing at least part of a gene necessary to maintain silencing of another gene in a cell or plant, comprising
  - (a) mutagenizing wild type cells or plants by randomly inserting into their genomes a DNA tag with known sequence;
  - (b) identifying mutants of said cells or plants which express RNA that is not expressed in wild type cells or plants;
  - (c) cloning a piece of genomic DNA surrounding or close to the insertion site of the DNA tag;
  - (d) screening a genomic library of wild type cells or plants with the piece of genomic DNA obtained in process step (c) or a part thereof;
  - (e) identifying clones comprising at least part of the gene affected by the insertion of the DNA tag; and
  - (f) further processing the clones obtained in step (e) using recombinant DNA techniques.
9. The method of claim 8, wherein plants are mutagenized using T-DNA, Ac/Ds or EN/I insertions, chemical or physical mutagenesis.
10. The method of claim 8, wherein mutants expressing RNA that is not expressed in wild type cells or plants are identified by reverse transcription of said RNA and subsequent amplification of the generated DNA using oligonucleotide primers specific for said DNA (RT-PCR).
11. The method according to claim 10, wherein pools of mutants are analyzed by RT-PCR.

12. The method according to claim 8, wherein the complete gene necessary to maintain silencing is produced.
13. A kit for the identification of plants impaired in transcriptional gene silencing comprising
  - a) a nucleic acid according to claim 1 conveniently labeled to be used as a hybridization probe or
  - b) an oligonucleotide primer for reverse transcription of RNA and an oligonucleotide primer specific for a nucleic acid according to claim 1.

*(085)*

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
22 March 2001 (22.03.2001)

PCT

(10) International Publication Number  
**WO 01/20010 A2**

(51) International Patent Classification<sup>7</sup>: C12N 15/82, (74) Agent: SYNGENTA PARTICIPATIONS AG; Patent and C07K 14/415, C12Q 1/68, A01H 5/00 Trademark Dept., Site Rosental, CH-4002 Basel (CH).

(21) International Application Number: PCT/EP00/08994 (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 14 September 2000 (14.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 9921964.4 16 September 1999 (16.09.1999) GB

(71) Applicant (*for all designated States except US*): SYNGENTA PARTICIPATIONS AG [CH/CH]; CH-4058 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): STEIMER, Andrea [CH/CH]; Solothurnerstrasse 92, CH-4053 Basel (CH). MITTELSTEN SCHEID, Ortrun [DE/CH]; Martin-skirchplatz 15, CH-4051 Basel (CH). PASZKOWSKI, Jerzy [PL/CH]; Blauenweg 10, CH-4224 Nenzlingen (CH).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/20010 A2

(54) Title: TRANSCRIPTIONALLY SILENCED PLANT GENES

(57) Abstract: The invention relates to gene silencing as observed after integration of transgenes into plant genomes. Comparison of transcriptional gene expression between an *Arabidopsis* line carrying a silent transgene present in multiple copies and its mutant derivative *mom1* impaired in silencing of the transgene revealed two cDNA clones which are expressed in the mutant plants, but not in the parental and not in wild type plants. Both clones are derived from the same family of transcripts referred to as TSI (Transcriptionally Silent Information). Genomic templates encoding TSI are repetitive elements with mainly pericentromeric location and conserved organization among various ecotypes. Transcriptional silencing of the genomic TSI templates is specifically released in the mutant. Transcription of TSI can be used as a marker to identify a defective silencing pathway in a plant.

**DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION** Original Supplemental Substitute

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

**TRANSCRIPTIONALLY SILENCED PLANT GENES**

the specification of which:

is attached hereto.

was filed on \_\_\_\_\_ as Application No. \_\_\_\_\_  
(day/month/year)

and, if this box () contains an x

was amended on \_\_\_\_\_  
(day/month/year)

was filed as Patent Cooperation Treaty international Application No.

PCT/EP00/08994 on 14/09/2000  
(day/month/year)

and, if this box () contains an x

entered the national stage in the United States and was accorded Application No.

and, if this box () contains an x

was amended, subsequent to entry into the national stage, on \_\_\_\_\_  
(day/month/year)

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment(s) referred to above and, if this application was filed as a Patent Cooperation Treaty international application, by any amendments made during the international stage (including any made under Patent Cooperation Treaty Rule 91, Article 19 and Article 34).

I acknowledge my duty to disclose all information which is known by me to be material to the patentability of this application as defined in 37 C.F.R. § 1.56.

I hereby claim the benefit under 35 U.S.C. §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate listed below and under 35 U.S.C. §365(a) of any Patent Cooperation Treaty international application(s) designating at least one country other than the United States listed below and have also listed below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application the priority of which is claimed for that subject matter:

COUNTRY/REGION (OR P.C.T.)	APPLICATION No.	FILING DATE (day/month/year)	PRIORITY CLAIMED	
GB	9921964.4	16/09/1999	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No
			<input type="checkbox"/> Yes	<input type="checkbox"/> No
			<input type="checkbox"/> Yes	<input type="checkbox"/> No
			<input type="checkbox"/> Yes	<input type="checkbox"/> No
			<input type="checkbox"/> Yes	<input type="checkbox"/> No

I hereby claim the benefit under 35 U.S.C. § 119 (e) of any United States provisional application(s) listed below:

APPLICATION NO.	FILING DATE (day/month/year)

I hereby claim the benefit under 35 U.S.C. §120 of any United States application(s) listed below and under 35 U.S.C. §365(c) of any Patent Cooperation Treaty international application(s) designating the United States listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of 35 U.S.C. §112, I acknowledge my duty to disclose all information known to me to be material to patentability as defined in 37 C.F.R. §1.56 which became available between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

United States Application No.	United States Filing Date (day/month/year)	Status (Pending, Abandoned or U.S. Patent No.)	International Application No. and Filing Date (day/month/year)

I hereby appoint the registered practitioners associated with Customer No. 022847, respectively and individually, as my attorneys and agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

If these brackets contain an X [X], I hereby authorize the registered practitioners associated with Customer No. 022847 and any others acting on my behalf to take any action relating to this application based on communications from the Patents and Trademarks Division of Syngenta Crop Protection AG, Basle, Switzerland, or an affiliate thereof or a successor thereto, without direct communication from me.

Please address all communications to the address associated with Customer No. 022847, which is currently Larry W. Stults, Syngenta, Patent and Trademark Department, 3054 Cornwallis Road, Research Triangle Park, NC 27709.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001 and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Full name of sole  
or first joint inventor *J-00*  
Andrea STEIMER

Inventor's signature

*Andrea Steimer*

Date

*06/02/2002*

(day/month/year)

Residence 8057 Zürich, Switzerland

*SEX*

Citizenship Swiss

Post Office Address Langackerstrasse 49  
8057 Zürich  
Switzerland

**IMPORTANT:** Before this declaration is signed, the patent application (the specification, the claims and this declaration) must be read and understood by each person signing it, and no changes may be made in the application after this declaration has been signed.



10/088384

## SEQUENCE LISTING

<110> Novartis AG  
 Novartis Research Foundation

<120> Transcriptionally Silenced Plant Genes

<130> S-31147A

<140>

<141>

<150> GB 9921964.4

<151> 1999-09-16

<160> 26

<170> PatentIn Ver. 2.1

<210> 1

<211> 2512

<212> DNA

<213> Arabidopsis thaliana

<400> 1

```

tcggcggtcc agcttcgcac actctctcac cgtctctcggt ttcactcgac cacttcacac 60
ttcgccctcaa catcttcgccc ggagtttctc gecattgtcc gtgcttccgt catctccgtt 120
caactcgacca ccggaccggc ttcaccatct ctcaactatac caccattcac tcgacacctcg 180
cattcactgc gcctccattc gtctctttac tcgactgctc ctcaaaaccgc caccgttcc 240
tctaaattcg ccgtttactc gaccacactg ttacgtctct cattcgtgtt cagtcgaccg 300
ctataccccga agccacaata tcactctact cgaccgtttc actcgatcgc gtacttgact 360
ggtttagtgt gtgtgtttat ttgaactaac atattgatat ttggtttga gttacattct 420
ttttcaggga atcaaatatga gcaactacag tggcgaatcc tccatggatg cggattacaa 480
cgtcgatgaa gctgaatctt ggtcaactag accagagaga gagcaacagg cttatgagag 540
cttcagagcc gagacccaac gtcagtagc tcgacgcaat gaaaggagag ctgagattgc 600
tagagggaaag agagcgatga ccagcagata ttagttgatc gacgaagata ttgacgatcga 660
gtatgagcct gagtcatggc acagagaaac aaaactgttg aacaaggctg atgaagtta 720
agtggaaagag tacatcagac ttttcgagct gaacgacttc tggggagcga ggtacccctg 780
ttatgagact ctggcccaac ttaggctact ggaggacgta cagcacttat tcgagaagtg 840
ccatcttgag acgctgtatgt cttaccgtt cgtcgcttac aagaaggaaa caatagagtt 900
tctctccact ctgcaagtgg agttgtatca gggacttact gcagatgaac tggagagtga 960
agggttggga ttcttgactt tttcagttaa cgacgcgtt taccagctat ctatcaagag 1020
cttggaaagga ttatgggtt ttcccagtgg aaagggaaact aaacccaagt tcgaaaaggga 1080
agagttgaag gatttgggt taaccattgg gaacgatttg ggcgctcaact ctgcaaggtc 1140
taagagcaac cagattcgaa gccctgtat cogctactat cagcgcctcg tagcgaatgt 1200
tctgtacccc agggaaatcta caggcaccgt gtctaacaca gacatggaga tgattgattc 1260
tgcactcaag ggtattctcc ggagaacaaa gggagaagaag gtcctaaagg ggcaccaa 1320
tgatacacca ccggcatgac ttctgtgtat ccataatgtgt ggatacagga agtgggcgca 1380
caccaacggg aggaagaagg tgcgaggage ctttgggtt ggtggcgat tgacaccgat 1440
tctgattgca tgggtgtac ctctcacgtc tccagggttt gatccgagga tgatggattt 1500
agatcacttg cgtcgttggg agtttctggg gtacgacatg gttggcgatt tctatcgcta 1560
caaattcgag cactccctga cccgaacacgc caacatggg cttccctgca tcgaggccac 1620
aaccatactt taggggtgaga acattgactt cagacctgctc cgtgattacc tctactttga 1680
gagcactcca ccgactgtatg acaatgtccc tacgacggaa gctacagagg atgatggatc 1740
tgagacggat gaggataggg aggaggagta tgatacggc atgtatcatt tcagtggatc 1800
cgtacctcca ggcgaggaga gcaagagctt gagtgaagct cacagaaaca acagtaagtt 1860

```

gcagagggtgg	tgcagaagaaac	aagataggtt	acttatcaag	tgcttcaagg	ccatcacgtt	1920
tctaacggac	aagataagt	gtttcttcc	taccacagct	attccgcagg	gagagcgccc	1980
tcaggacatg	ccttcgaaga	gatatgacgc	gccaggccca	agtcatcaca	ggcctgagcc	2040
aagtccaccac	aggcctgagc	ctagtgaccg	agttagtacca	ccagttccctg	caaggcattc	2100
atcattcgag	cctcgggagc	tcggggagaaa	gaagaaggct	gcactcgcta	ggtctggcag	2160
caggagtaga	cgacttctcc	agtcccgtag	cttacgcgac	cgcggtgctg	gcccgcagcag	2220
aagaagagag	gtcgagtatc	atcagagcgg	tgctggccgc	ggcgaaggag	cagaggtcga	2280
gtaccccccag	gggaaagctg	agacacaaca	gggagattct	tcgatggcct	gggagcaatc	2340
acagggcagct	attgacgacc	aactccgctc	cttcttccac	tgaggtatgc	acctcactcc	2400
accattgttaa	tataccatct	cttgtttttt	attttgtttt	tgtgtatgtgt	tttgtccctga	2460
gtactcttt	ccaaaatttgg	tcacacagtg	gactgtgtga	tttaagtttg	gg	2512

<210> 2  
<211> 1997  
<212> DNA  
<213> *Arabidopsis thaliana*

<400>	2					
ccgggcagg	caacaggctt	atgagagctt	tagagcttag	accacaacgt	cagtagctcg	60
acgcaatgaa	aggagagctg	agattgctag	agggaaagaga	gcaatgacca	gcagatata	120
gttgcgtac	gaagatattg	acgtcagact	tgaacctgaa	tcatggcaca	gagaaaacgaa	180
gctgttgaac	aagcccgatg	aagttacagt	agaggagtag	atcagacttt	tcgagctgaa	240
cgacttctag	ggaacgaggt	accctgtta	tgagacttta	gcccagctgg	gcctactgaa	300
ggacgtacat	cacttattcg	agaagtgc	tctggagacg	ctgatgttct	accggtaact	360
cgcttacaag	aaggaaacaa	tagagttct	ctccactctg	caagtgagaa	tgtatcaggg	420
acttactgca	gatgagctgg	agagtgaagg	gttggggttc	ttgacttttt	cagttaacga	480
gcagcgttac	cagctatcta	tcaagagact	ggaaggattt	tttgggtttc	caagtggaaa	540
ggaaactaaa	cccaagttcg	agagggaaaga	gttggaaagat	ttgtgtttaa	ccattgggaa	600
cgatggca	ctcaactctg	caaggctcaa	gagcaaccag	attcgaaagcc	ctgtgatccg	660
ctactatcg	cgctcagtag	cgaatgttct	gtaccccagg	aatctacag	gcacgtgtc	720
taacacagac	atggagatga	ttgatgttgc	actcaaggggc	attctccgga	aaacaaagggg	780
gaagaaggtc	ctaaaggccg	accttaatga	tacaccaccc	tttatgttcc	tgttgatcca	840
cctgtgtgga	tacaggaagt	gggcgcacac	caacgagaag	aagaaggc	gaggagccct	900
tttgttaggt	ggcgttgtga	caccgattct	gattgcattgt	gtgtgtacctc	tcacgtctcc	960
agggtttgtat	ccgaggatga	tggatttaga	tcacttgcgt	cgttgcgt	ttcttagagta	1020
cgacatggtt	ggcgttct	atcgctacaa	attcgagcac	tccctgaccc	gaacagccaa	1080
cattttgtt	ccctgcac	aggccacaac	cataacttcg	gttgagaaca	ttgacttcag	1140
acctgcgcgt	gattacctct	actttgagag	cgctccaccc	actgtgaca	atgtccctac	1200
gacggaaagt	acagaggatg	atattgctga	gacggatgag	gatagggagg	aggagatata	1260
tacgagcatg	tatcatttca	gtgagcacgt	acctccacgc	cgggagagca	agagctttag	1320
tgaagctcac	agaaaacaaca	gtaaatgtca	gagggtgtgc	aagaacaac	ataggctact	1380
tatcaagtgc	ttcaaagcca	tcacgttct	gacggacaag	ataagctgt	tctcttctac	1440
cacagctatt	ccgcaggggag	agcatcctca	ggacatgcct	tcaaggagat	atgacgcgcc	1500
agcgccaagt	catcacaggc	ctgagccaag	tcaccacagg	cctgagccta	gtgaccgagt	1560
agtcccccacca	gtccccgtcaa	ggcattcatc	attcgacgcct	cgggagctcg	ggagaaaagaa	1620
gaaggctgca	ctcgctcggt	ctggcagcag	gagtacacga	tttctacagt	cccgtagctt	1680
acgcgaccgt	ggtgtctggcc	gcagcagaag	aagagaggc	gagtatcatc	agagcgggtc	1740
tggccogcgac	gaaggaggcag	aggtcgagta	ccccacgggg	aagctgagac	acaacaggga	1800
gattcttcga	tggcctggga	gcaatcacat	gcagctattt	acgaccaact	ccgctccctc	1860
ttccactgag	gtatgcacct	cactccacca	ttgtatata	ccatctcttgc	tttttatttt	1920
gtttttgtga	tgtgttttgt	cctgagact	ctcttccaaa	tttggtcaca	cagtggactg	1980
tgtgatttaa	gtttgggg					1997

<210> 3

&lt;211&gt; 1682

&lt;212&gt; DNA

<213> *Arabidopsis thaliana*

&lt;400&gt; 3

cttatattat gggttggat gtgttaaag aaaaggggaa attcattgtt gataaggaaa 60  
 gggaaagaat tctagggaa gtaagctaaa gaagttagaa aaaatctagt aaaggtttg 120  
 ggaatgttaa agaaaagaat gaggttctt gtagctaaag aataagggtt aaaagcctt 180  
 ggtttaaag attaaaaaaaaaa aaacaggAAC cttagttgtt aaagaatcc aaaccgccta 240  
 gatgtatcaa gagcgTTGAG aaagcttctc cttagttaa gagaaaagaa aagaatgata 300  
 tgaaaaaagag ttgaaaagat tcatgagtgc aaaggtaga gttaaaggtagg gacaggagtt 360  
 ggttttacca tttagaacttc attgttatac tctgggtaga tgggatctta tctctgtatg 420  
 cataatttgg gacttacctt tagcattcta ctaaagctca atcatttttg aggatcccc 480  
 tgttacttaa gccttattcta taagggacca tctttgtctc ttgaccttca cttggccga 540  
 atgagttcat tgatgatgca ttgctttagt cgcgTTCCAG aactaatgaa ttgaaaagg 600  
 attggtagat ttgaaagcat gtgttaggtcg agtataagag acggattgtatg tggaaaacaag 660  
 gcatggctaa cgtttttagg tagaattcaa tcatatcgca tctttagaact accaacttgg 720  
 acattgattt tatttgcctt atcatatgct ttggTTTGA gtcccccgcct tcaactcctct 780  
 ccttcaacta tgcatttctt ttgctttagt ggcaagcAAA gactaagttt gaggagttg 840  
 atatgtctat aatttgcattt ttgcattgtt ccattcatca tgcatttttagg tccagttcg 900  
 tatcattcat cactgtttta tatcatttctt catcattctt gcatactttg catgatttagg 960  
 ataacttgc atacatattt catttcttagt ttgttttagt gtgatttgg gctgtttgca 1020  
 agcaaattgg aagaaatgag ccagaaccAG aagacatact cgacccttag gtcgagtgc 1080  
 ttggggcca ttcttccac atactcgGCC cccaggTCGA gtgactttgg accattttt 1140  
 cccatccact cgaccaccGG gtcgagtaac cttagcttagt gcaactcgat gacactactc 1200  
 gaccccaagt cgagtatcac ttgcacac cacctgacAA cactcgacca atcaacttac 1260  
 caagttactc gaccccttgg tcgagtatca tcaactcacca ccatcagcat cactcgaccg 1320  
 gacactcgat cacgttctca cagtctactc aaatccgcAG tcaaccagAC aagctgagca 1380  
 caaggaagag aagaggagaa gacaaagtgc ttggaagcgg cctggacctc catcgatca 1440  
 cgaagcccat ctggcccat tatctctctt tggccgagc gattaggtt ttggccgc 1500  
 tactatcatt ttatTTGTT ttgtataaat agatgtcttA gggTTTGTc ctgagacatc 1560  
 tagtcgacat tgatTTTTT ttgcTTTCAgT ttatTTTCT gttctactct gtcgcgc 1620  
 ttgtcttctt gcaacctgta attcgagatt ttccaagttt attcagattc cgcatTTGAT 1680  
 tt

&lt;210&gt; 4

&lt;211&gt; 1652

&lt;212&gt; DNA

<213> *Arabidopsis thaliana*

&lt;400&gt; 4

cttatatcat gggTTGGAT cagttaaaa aaaaaaaaaAGG gtgaattcat ttgtgataag 60  
 gaaaggggaaa gaattctagg ggaagtaAGC taaagaAGTt agaaaaaaaaaa aaatcttagta 120  
 aaggTTTGG gaatgttaaa gaaaAGAATg aggttcttGT tagctaaAGA agaagggtt 180  
 aaagcctttt ttgttaaAGA ttAAAAACAG gAACCTTAGt ttgttaaAGAA atccaaatac 240  
 gctagatgtA tcagagtGTT gagaAGCTT ctccTAGAGt taagAGAAA gaaaAGAATG 300  
 atatgaaaaa gagTTGAAA gattcatgag tgcaAAAGGGT agagttaAGT tcttGTATTG 360  
 ggactggagt tggattacc attAGAGCTT cattgttATA ctatggtagt atgggatttt 420  
 atctctgtat gcataacttG ggacttACCT ttagcattct actaaAGCTC aatcattctt 480  
 gagagatccc ctgttacttA agcctattct gtaaggGACC atctttgtct ctgaccttc 540  
 accttagcca aatgagttca ttgatgatgc attgtttgtat tcacgttcca gaactaatga 600  
 atgttaaagg gattggtagA ttgaaAGCA tggtaggtc gaggataAGA gacggattGA 660  
 ttgataacAA ggcattggcTA acgtttcGA gtaAAATTCA atcatatcgC atcttagaaAC 720  
 taccaacttG gacattgatt ttatTTGCTC tattcagatgc ttggTTCTG agtccccacc 780  
 ttcaaacctc tccttcaact atgtcttctt atttgcttGA gggcaagCAA agactaAGT 840

tggggaggtt	gatatgtcta	taatttgcattttcagtg	tccattcatc	atcgaaaaat	900	
gtccaggat	tttcgtatcattca	tcactgtttt	atatcatttc	tcatcattct	tgcatacttt	960
gcatgattt	tagatgttttgc	tatacatatt	gcatttctga	gttgtttta	ggtgatattgg	1020
agctgttttgc	gagcaaaatttgc	gaagaaacga	gccagaacaa	gaagcatac	tcgaccaccc	1080
ggtcgagt	tttggagcc	attcttccca	tctactcgac	ccgggggtcg	agtaaacctca	1140
gctcagg	ccctcgatgacg	ccactcg	ccctggtcg	agtatca	ctccacacca	1200
cctgacc	cactcttgcgtt	cactcttacca	cgttactcga	ccccctggtc	gagtatcatc	1260
actcacc	aaccatca	ctcgacccggg	cactcgatca	catcttcata	gtctactcaa	1320
atccgcact	aaccagacaa	gctgagcaca	aggaagagaa	gaggagaaga	caaagtgttt	1380
ggaagcgg	cccttggac	tcggatcacg	aagaagccca	tctcgccca	ttatcattct	1440
atgggc	cgat	atggccgt	ctactatcat	tttatttgcgt	tatgtataaa	1500
tagatgt	tttgcgtt	accaggacat	ctagtcgaca	ttgagttttt	ttgcttcag	1560
tttat	tttgcgtt	ctgcgcgcgt	tttgcgttgc	caacctgtaa	ttcgagattt	1620
ttccaagtt	ctcagattcc	gcatttgcattt	tc			1652

<210> 5  
<211> 903  
<212> DNA  
<213> *Arabidopsis thaliana*

<210> 6  
<211> 614  
<212> DNA  
<213> *Arabidopsis thaliana*

<400> 6  
gagggcaagg aaagactaaag tttgggggag ttgataagtgc tgtatTTTgc atgtttttag 60  
catccatttg tcatcacttt agcatcatat catcaCTGTT ttataccatt tcacatcatt 120  
tgtcatcaact ttgcATGTTT aggatagttt tgcatgcATG ttgcATATT gcgttGATT 180  
caggtgattt ggAGCTGTTG acgagctatC tggAAAGAAGC agacCTGATC atgacaAAAC 240  
actcgaccca gaggtcgagt aggAGCTCA agatCTCAAG agactactCG acaACCAGGT 300  
cgagtagAGC acatCACCCAC ttcACCTCAC cactcgACCC CGAGTCGAG tgccatCATC 360  
tccatcacca gacGGTCACT cgatCACTC actcgacCTT gaggtcgAGT gtcttcACCT 420  
ccatcatcag acaACCACTC gacCTCCTCA ctccACCTAG aggtcgAGTA tctccatCTT 480  
accactcgcac tgcataACTCG atgacaAGCT tcAGAGCCTT cttAAATTCCG cactcaACCA 540  
gacactcgcAG cacaAGGAAG AAAAGAAGAC tccAGCTATT cactcgAGCT ctcactcgAC 600

cacgtgggtc gagt

614

&lt;210&gt; 7

&lt;211&gt; 1956

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis thaliana

&lt;400&gt; 7

ttttggttca ccggatttaac tcagtgcctc cggtgctagt tgggtgttgc gtagtgaatg 60  
 aatttggaaag aaaactgaac catgcctaga ttgcctact cgaccacact gtcatgatct 120  
 gataccatc cctatcaatt tgaacctgaa tttgatcttt aattatcatg tctgcattcaa 180  
 atttgaactc atggataccct taaaataactt ggattttctt attcattttg atcactcttg 240  
 ttaatccaag tagctgactc tccttattag agcagttAAC ccgaacccaa acctagactt 300  
 ttttcaagg ctatatacac tgcgtgggt ttgtgaggTC ttattccat tcagcttgg 360  
 agaaagtgtt aggtctgtAA cgacagagat agtgnctcat gtagttctag ttgcatttt 420  
 ttggactaga taggactggg tgggcgttA tacttttagt tggtatngt taaaagaaaa 480  
 aaaaaaggggg ttgattcatt gatgagaaaa ggtaaaagac tctaggtAA gtaagataaa 540  
 gaagcagaaaa aggtctgtAA aaggTTTGG gatttgtAAA aaaaagaaaag agttcttgg 600  
 agctattgaa gatggggAAA agccctcggt tttaaatgt taaaaacagg aaccttagtt 660  
 gttaaagaaaa tccaaatccg ctagatgtat caaatgttt agaaagctt tcttagagtt 720  
 aagagaaaaag aaaagaatga tttagaaaaAG ggcttaaagg attcatgat gcaaagggtA 780  
 gaggttaattt ctatatactgg gattggagat gggattacca tttagacttc atctgatata 840  
 ctcttaggtat atggatctt atctctgtat gcatagtttt ggacttacct ttgcattct 900  
 actaaagctt aatcatTTTT tgagagatcc cctgttactg aagccttattc tgaaggggac 960  
 catcttgc tcttgacctt ttaccttagc caaatgagtt cattgtat gcatgttgc 1020  
 attcacgttc cagaactaat gaatgtAAA gggattggta gatttgaaaa catgtgttag 1080  
 tcgagcatat gatcggtt gattgtatgt aaggcatggc taaagtTTT cagtagaatt 1140  
 cgatcatatc gcagcttagA actatcaact tggacattga ttcatTTGG ttatcttagt 1200  
 gcttggctc tgagtccccg atttcaaac tcacctctag ctgttctta attgtttgct 1260  
 tgagggcaag caaaactaa gtttggggA gttgataagt gtgtatTTG catgtttga 1320  
 gcatccattt gtcatcactt tagcaccata tcatactat ttataccat ttctcatcat 1380  
 ttgtcatcac ttgcattgtt tagatagtt ttgcattgtat gtggcatatt tgggtgttt 1440  
 tcaagtgtt cggagctgtt gaagaactaa ttggaaAGC cggacctgtat catgccaaac 1500  
 cactcgacct caggtcgagt agacgcttca cgacctaAC acaccactcg accacctgg 1560  
 cgagtgttagg acttcaccac ttccacccat cactcgaccc cctggccgag taccacca 1620  
 gagtcactcg atcaattcac tggacccca ggtcgagtgtt ctccacccat accacctgac 1680  
 catcaactcgA tcacacgact ctacctggAA gtcgagtatc accatcacca ccactcgact 1740  
 acataacttga tgcgtgggtt cagacttcc tccattccgc actcaaccag acactcgagc 1800  
 acaaggaaaa aaagaagatt ctatctatc actcgacctc tcactcgacc acctgggtcg 1860  
 agtacagttc ttaatccgtc tcaatactgc gtcgtttga gtattagggt ttccgaatat 1920  
 tttgtata agtagcacgt actttacatt ttccgag 1956

&lt;210&gt; 8

&lt;211&gt; 2105

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis thaliana

&lt;400&gt; 8

ttagcatttt ggttcacttag ataaaactcag tggccctcggtt gttagttgtc tgcatac 60  
 tcaatggaaat tgaagtaaaa ctgcaccatg cctagattgc tctactcgac cacactgtta 120  
 ggatctgata ccattcccta tcaatttgAA ctgttatctg attttagaaatt atcatgtctt 180  
 gccatcgaaat ttgaactcat ggataccctA aaatacttgg attttcttac tcattttaa 240  
 cactcttggttt aatccaagta gctgactctc ctatttagag cagttAAC gaatccaaac 300  
 ctaatcttcc ttgcgtggccc tatatcactt gtgagttttt gtgaggtctt atttcaattt 360

agcttggtag aaagtgttag gttcgtaacg acagagatag tgtctcatgt agttctagtt 420  
 cgcgttttt ggactggata ggacttaggtg ggcgcttata tcatggttt ggatcagtt 480  
 aaaaaaaaaa aagggtgaat tcattgtta taaggaaagg gaaagaattc tagggaaagt 540  
 aagctaaaga agttagaaaa aaaaaaatct agtaaagggt ttggaaatgt taaagaaaag 600  
 aatgagggttc ttgttagcta aagaagaagg gttaaaagcc ttttgtttt aagattaaaa 660  
 acaggaacct tagttgttaa agaaatccaa atacgctaga tgtatcagag tggtgagaaa 720  
 gcttcctcta gagtaagag aaaagaaaag aatgatatga aaaagagttt gaaagattc 780  
 tgagtcaaaa gggtagagtt aagttcttgtt attgggactg gagttggat taccattaga 840  
 gttcattgt tatactatgg gttagatggg ttttatctct gtatgcataa cttgggactt 900  
 accttagca ttctactaaa gctcaatcat tcttgagaga tcccctgtta ctaagccta 960  
 ttctgttaagg gaccatctt gtctcttgcc ttcacccctt ccaaattgagt tcattgtat 1020  
 tgcattgttt gattcacgtt ccagaactaa tgaatgttaa agggattggg agatttgaaa 1080  
 gcatgtgttag gtcgagtata agagacggat tgattgataa caaggcatgg ctaacgtttt 1140  
 cgagtaaaat tcaatcatat cgcattttt aactaccaac ttggacattt attttatttt 1200  
 ctctatcaga tgctttgggtt ctgagttcccc accttcaaac ctctccttca actatgtctt 1260  
 cttatttgct tgagggcaag caaagactaa gtttggggg gttgatatgt ctataatttt 1320  
 catgttttca gtgtccattt atcatgttt tgagtccagt ttcgtatcat tcattactgt 1380  
 tttatatcat ttctcatcat tcttgatatac tttgcatgat taggatagct ttgtatacat 1440  
 attgcatttc tgagttgttt ttaggtgatt tgagctgtt tgcgagcaaa ttgaaagaaa 1500  
 cgagccagaa caagaagcca tactcgaccc cctggctcgag tgactttgga gccatttttc 1560  
 ccatctactc gaccgggggg tcgagtaacc toagctcagg ccactcgatg acgcccactcg 1620  
 tccccctgg tcgagtatca ctccgcaca ccaacctgacc acactggcc gttcaactcta 1680  
 ccacgttact cgacccttgc gtcgagttatc attactcacc accaacacca tcactcgacc 1740  
 gggcactcga tcacatctt atagtcact caaatccgca ctcaaccaga caagctgagc 1800  
 acaaggaaga gaagaggaga agacaagtg ttggaaagcg gcctggaccc ccatcgatc 1860  
 acgaagaagc ccatctcgcc ccattatcat tctatggcc gggcgattag gttattggcc 1920  
 cgtctactat cattttatcc cgttatgtat aaatagatgt tttaggttcc ttgtaccagga 1980  
 catctagtcg acattgagtt ttttgcctc agtttattt tctgtttct ctgctgcgcc 2040  
 gctttgttt ctgcaacctg taattcgaga ttttccaag ttattcagat tccgcattt 2100  
 atttc 2105

<210> 9  
 <211> 4860  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Virtual TSI

<400> 9  
 tcggcggtcc agcttcgcac actctctcac cgtctctcggt ttcactcgac cacttcacac 60  
 ttccgcctcaa catcttcgccc ggagtttctc gccattgtcc gtgctccgtt catctccgtt 120  
 cactcgacca cgggaccggc ttccacatct ctcactatcc caccattcac tcgacccctgc 180  
 cattcaactgc gcctccatttgc ttctctttac tcgactgctc ctcaaaaccgc caccgtcttc 240  
 tctaaattcg ccgttactc gaccacactg ttacgtcttot cattcggttta cagtcgaccc 300  
 ctataccga agccacaata tcacttcact cgaccgtttc actcgatcgc gtacttgact 360  
 ggtttagtgtt gtgtgtttat ttgaactaac atattgatat ttggtttga gttacattt 420  
 ttttcagggaa atcaatatga gcaactacag tggcgaatcc tccatggatg cggattacaa 480  
 cgtcgatgaa gctgaatctt ggtcaacttag accagagaga gagcaacagg cttatgagag 540  
 ctccagagcc gagacccaaac gctcagtagc tcgacgcaat gaaaggagag ctgagattgc 600  
 tagagggaaag agagcgatga ccagcagata tgagttgatc gacgaagata ttgacgtcga 660  
 gtatgagccct gaggcatggc acagagaaac aaaactgtt aacaagcctg atgaagttac 720  
 agtggaaagag tacatcagac ttttcgagct gaacgacttc tggggagcga ggtacccctg 780  
 ttatgagact cttagccccago tttaggttact ggaggacgtt cagcacttat tcgagaagtg 840  
 ccatcttgatc acgctgtatgtt cttaccgttca cgtcgcttac aagaagaaa caatagagg 900

tctctccact ctgcaagtgg agttgtatca gggacttact gcagatgaac tggagagtga 960  
agggttggga ttcttgactt tttcagttaa cgagcagcg taccagctat ctatcaagag 1020  
cttggaaagga ttatttgggt ttcccagtgg aaagggaaact aaacccaagt tcgaaaggga 1080  
agagttgaag gatttgggt taaccattgg gAACGATTG gCGCTCAACT CTGCAAGGG 1140  
taagagcaac cagattcgaa gccctgtgt cgctactat cAGCCTGAG TAGCAGATGT 1200  
tctgtacccc aggaatcta cAGGCACCGT GTCTAACACA GACATGGAGA TGATTGATT 1260  
tgcactcaag ggtattctcc ggagaacaaa ggggaagaag gTCCTAAAGG GCGACCTAA 1320  
tgatacacca cgggtcatgc ttctgttgat ccataatgtgt ggatacaggag AGTGGGCGCA 1380  
caccaacggg aggaagaagg tgcgaggagc cttttgtgt ggtgggttg tgacacggat 1440  
tctgattgca tgggtgtac ctctcacgtc tocagggttt gatccgagga tgatggattt 1500  
agatcacttg cgtcggtgt agtttggta gtacgacatg gttggcgatt tctatcgcta 1560  
caaattcgag cactccctga cccgaacagc caacatTTT CTTCCCTGCA TCGAGGCCAC 1620  
aaccatactt tagggtgaga acattgactt cagacctgcg cgtgattacc tctactttga 1680  
gagcaactcca ccgactgtat acaatgtccc tacgacggaa gctacagagg atgattttgc 1740  
tgagacggat gaggataggg aggaggagta tgatacggc atgtatcatt tcagtgagca 1800  
cgtacctcca ggcaggaga gcaagagctt gagtgaagct cacagaaaaca acagtaagtt 1860  
gcagagggtgg tgcaagaaac aagataggtt acttatcaag tgcttcaagg ccatcacgtt 1920  
tctaacggac aagataagtt gcttcttcc taccacagct attccgcagg gagagcgtcc 1980  
tcaggacatg ctttgcaga gatatgacgc gccaggggca agtcatcaca ggcctgagcc 2040  
aagtccaccac aggcttgaggc ctatgtaccc agtagtacca ccagttccctg caaggcattc 2100  
atcattcgag cctcgggagc tcgggagaaa gaagaaggct gcactcgcta ggtctggcag 2160  
caggagtaga cgacttctcc agtcccgtag ctacgcgac cgcggtgctg gccgcagcag 2220  
aagaagagag gtcgagttac atcagagccg tgctggccgc ggcgaaggag cagaggtcga 2280  
gtaccccccag ggggaagctg agacacaaca gggagattct tcgatggcct gggagcaatc 2340  
acaggcagct attgacgacc aactccgctc cttcttccac tgaggtatgc acctcactcc 2400  
accattgtaa tataccatct cttgtttttt atttgtttt tigtatgtgt ttgtcttga 2460  
gtactctctt ccaaatttgg tcacacagtg gactgtgtga tttaaagttt ggggagggct 2520  
caggaagtgt gtgtgcatt gtataatctt gagtttgcatt tcatctaagg catagaaaaaa 2580  
ccaaaaaaaaat tgaaaaattc cagaaaaatga tttcacaaaaa atagagtgtt catgtagttg 2640  
cattgcattt aggatcgagt ttagagtgtt tggtttagga ttgttcata tgcatagggg 2700  
ataataatga gatagcctt gtaagcatttt gtttcaaccatg ataagcttag cgcctcggt 2760  
gttagttgtt tgatgcgtt tcattgaaat tgaagtaaga actgcacgat gccttagattt 2820  
ctctactcga ccacactgtt aggatctgtat atcattccct atcaatttga acttgaatct 2880  
gatttagaat tatcatgtct tggcatcgaa ttgaactca tggataccct aaaatactt 2940  
gattttctta ctcattttaa ccactctgt ttagtccaagg agctgactct ctttattttt 3000  
gcagtttaacc cataccccaaa cctgaacttt ctttcaagcc ctatatact tggatgttt 3060  
tgtgaggctt tatttcgatt gagcttggta gaaaagtgtt ggttgcata acagagata 3120  
gtgtctcatg tagttctgt ttcgttctt cagactggat aggacttagt gggcgcttat 3180  
atcatgggtt gggatgtgtt taaaagaaaa ggggaatctt atgttgatg agggaaaggga 3240  
aagaattcca ggggaagttaa gctaaagaag ttagaaaaaaa aatctgttagt aggttttggg 3300  
aatgttaaaag aaaagaatga gtttcttggt agctaaagaaa gaagggtttaa aagcccttgg 3360  
ttttaaagat taaaaaaaaaa acaggaacct tagttgttaa agaaatccaa acccgctaga 3420  
tgtatcaaga gcttgagaa agtttctct agagttaaaga gaaaagaaaa gaatgatatg 3480  
aaaaagagtt taaaagattc atgagtgcac agggtagagt taagtggga caggagttgg 3540  
tttaccatt agaacttcat tgtttactc tgggttagatg ggtatcttac tctgtatgca 3600  
taatttggga ctaccttta gcattctact aaagctcaat cattctttag ggtatcccctg 3660  
ttacttaagc ctattctata agggaccatc ttgtctctt gaccccttacc ttggccgaat 3720  
gagttcattt atgatgcatt gcttgattcg cggtccagaa ctaatgaatg taaaagggat 3780  
tggtagattt gaaagcatgt ttaggtcgag tataagagac ggattttagt aaaacaaggc 3840  
atggcttaacg ttttgagta gaattcaatc atatcgcatc tttagaactac caacttggac 3900  
atggattttt ttgtctctat catatgtttt ggttttgagt ccccgcttc actcctctcc 3960  
ttcaactatg tcttcttatt tgcttgaggg caagcaaaga ctaagttgtt gggagttgtt 4020  
atgtctataa ttgcgtgtt ttcatgtcc attcatcatc gtttttagtc cagtttcgtt 4080  
tcattcatca ctgttttata tcatttctca tcatttcttc atactttgca tgatttagat 4140  
aactttgcat acatattgca tttctgagtt gttttcaggt gatttggagc tgtttgcag 4200  
caaattggaa gaaatgagcc agaaccagaa gacatactcg acccccttaggt cgagtqactt 4260

tggggccatt cttcccacat actcgcccc caggtcgagt gactttggag ccattcttcc 4320  
catccactcg accaccgggt cgagtaacct tagctcaggc cactcgatga cactactcg 4380  
ccccaggctcg agtatactt cggccacacca cctgacaaca ctgcaccaat cactctacca 4440  
agttactcg cccccctggtc gagtatcatc actcaccacc atcagcatca ctcgaccgg 4500  
cactcgatca cgtcttcaca gtctactcaa atccgcagtc aaccagacaa gctgagcaca 4560  
aggaagagaa gaggagaaga caaatgtgtt ggaagcggcc tggacctcca tcggatcag 4620  
aagcccatct cggcccatta tctctctatg ggccgagcga ttaggttatt ggcggctca 4680  
ctatcattt atttcgtttt gtataaaatag atgtcttagg gttttgtcct gagacatcta 4740  
gtcgacattg agttttttt gtttcagttt tattttctgt tctactctgc tgcccgctt 4800  
ttgcttctgc aacctgtaat tcgagatttt tccaagttat tcagattccg catttgattt 4860

&lt;210&gt; 10

&lt;211&gt; 648

&lt;212&gt; PRT

&lt;213&gt; Arabidopsis thaliana

&lt;400&gt; 10

Met	Ser	Asn	Tyr	Ser	Gly	Glu	Ser	Ser	Met	Asp	Ala	Asp	Tyr	Asn	Val
1									10					15	

Asp	Glu	Ala	Glu	Ser	Trp	Ser	Thr	Arg	Pro	Glu	Arg	Glu	Gln	Gln	Ala
														30	
20										25					

Tyr	Glu	Ser	Phe	Arg	Ala	Glu	Thr	Gln	Arg	Ser	Val	Ala	Arg	Arg	Asn
35									40				45		

Glu	Arg	Arg	Ala	Glu	Ile	Ala	Arg	Gly	Lys	Arg	Ala	Met	Thr	Ser	Arg
50									55			60			

Tyr	Glu	Leu	Ile	Asp	Glu	Asp	Ile	Asp	Val	Glu	Tyr	Glu	Pro	Glu	Ser
65									70			75		80	

Trp	His	Arg	Glu	Thr	Lys	Leu	Leu	Asn	Lys	Pro	Asp	Glu	Val	Thr	Val
85									90			95			

Glu	Glu	Tyr	Ile	Arg	Leu	Phe	Glu	Leu	Asn	Asp	Phe	Trp	Gly	Ala	Arg
100									105			110			

Tyr	Pro	Cys	Tyr	Glu	Thr	Leu	Ala	Gln	Leu	Arg	Leu	Leu	Glu	Asp	Val
115								120			125				

Gln	His	Leu	Phe	Glu	Lys	Cys	His	Leu	Glu	Thr	Leu	Met	Ser	Tyr	Pro
130								135			140				

Tyr	Val	Ala	Tyr	Lys	Glu	Thr	Ile	Glu	Phe	Leu	Ser	Thr	Leu	Gln	
145								150			155		160		

Val	Glu	Leu	Tyr	Gln	Gly	Leu	Thr	Ala	Asp	Glu	Leu	Glu	Ser	Glu	Gly
165								170			175				

Leu	Gly	Phe	Leu	Thr	Phe	Ser	Val	Asn	Glu	Gln	Arg	Tyr	Gln	Leu	Ser
180								185			190				

Ile	Lys	Ser	Leu	Glu	Gly	Leu	Phe	Gly	Phe	Pro	Ser	Gly	Lys	Gly	Thr
195								200			205				

Lys Pro Lys Phe Glu Arg Glu Glu Leu Lys Asp Leu Trp Leu Thr Ile  
210 215 220

Gly Asn Asp Leu Ala Leu Asn Ser Ala Arg Ser Lys Ser Asn Gln Ile  
225 230 235 240

Arg Ser Pro Val Ile Arg Tyr Tyr Gln Arg Ser Val Ala Asn Val Leu  
245 250 255

Tyr Pro Arg Glu Ser Thr Gly Thr Val Ser Asn Thr Asp Met Glu Met  
260 265 270

Ile Asp Ser Ala Leu Lys Gly Ile Leu Arg Arg Thr Lys Gly Lys Lys  
275 280 285

Val Leu Lys Gly Asp Leu Asn Asp Thr Pro Pro Val Met Leu Leu Leu  
290 295 300

Ile His Met Cys Gly Tyr Arg Lys Trp Ala His Thr Asn Gly Arg Lys  
305 310 315 320

Lys Val Arg Gly Ala Leu Cys Val Gly Gly Val Val Thr Pro Ile Leu  
325 330 335

Ile Ala Cys Gly Val Pro Leu Thr Ser Pro Gly Phe Asp Pro Arg Met  
340 345 350

Met Asp Leu Asp His Leu Arg Arg Cys Glu Phe Leu Glu Tyr Asp Met  
355 360 365

Val Gly Asp Phe Tyr Arg Tyr Lys Phe Glu His Ser Leu Thr Arg Thr  
370 375 380

Ala Asn Ile Leu Leu Pro Cys Ile Glu Ala Thr Thr Ile Leu Xaa Gly  
385 390 395 400

Glu Asn Ile Asp Phe Arg Pro Ala Arg Asp Tyr Leu Tyr Phe Glu Ser  
405 410 415

Thr Pro Pro Thr Asp Asp Asn Val Pro Thr Thr Glu Ala Thr Glu Asp  
420 425 430

Asp Phe Ala Glu Thr Asp Glu Asp Arg Glu Glu Glu Tyr Asp Thr Ser  
435 440 445

Met Tyr His Phe Ser Glu His Val Pro Pro Ala Gln Glu Ser Lys Ser  
450 455 460

Leu Ser Glu Ala His Arg Asn Asn Ser Lys Leu Gln Arg Trp Cys Lys  
465 470 475 480

Lys Gln Asp Arg Leu Leu Ile Lys Cys Phe Lys Ala Ile Thr Phe Leu  
485 490 495

Thr Asp Lys Ile Ser Cys Phe Ser Ser Thr Thr Ala Ile Pro Gln Gly

500	505	510
Glu Arg Pro Gln Asp Met Pro Ser Lys Arg Tyr Asp Ala Pro Gly Pro		
515	520	525
Ser His His Arg Pro Glu Pro Ser His His Arg Pro Glu Pro Ser Asp		
530	535	540
Arg Val Val Pro Pro Val Pro Ala Arg His Ser Ser Phe Glu Pro Arg		
545	550	555
Glu Leu Gly Arg Lys Lys Ala Ala Leu Ala Arg Ser Gly Ser Arg		
565	570	575
Ser Arg Arg Leu Leu Gln Ser Arg Ser Leu Arg Asp Arg Gly Ala Gly		
580	585	590
Arg Ser Arg Arg Arg Glu Val Glu Tyr His Gln Ser Gly Ala Gly Arg		
595	600	605
Gly Glu Gly Ala Glu Val Glu Tyr Pro Gln Gly Glu Ala Glu Thr Gln		
610	615	620
Gln Gly Asp Ser Ser Met Ala Trp Glu Gln Ser Gln Ala Ala Ile Asp		
625	630	635
Asp Gln Leu Arg Ser Phe Phe His		
645		

<210> 11

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 11

tggttcacca gataagctca gtgccctc

28

<210> 12

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 12

cttcagactg gataggacta ggtggcg

28

<210> 13  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 13  
cgccccaccta gtcctatcca gtctgaag

28

<210> 14  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 14  
cgcataacaac aactaacaac gagggcac

28

<210> 15  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 15  
cgataaacatc gaccgtattt ctcgcc

26

<210> 16  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 16  
aacttagctcc catccgtctt cgacatcc

28

<210> 17  
<211> 24  
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 17

tgcacatcacac cgggattggat tgac

24

<210> 18

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 18

tgttcccttg aaccatagca atgagacc

28

<210> 19

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 19

caaacagaca gagtgtggcc caccacc

27

<210> 20

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 20

caaacagaca gagtgtggcc caccacc

27

<210> 21

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

## Oligonucleotide sequence

<400> 21  
tgcaaaccctt caggaccaag tctaccc

27

<210> 22  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide sequence

<400> 22  
acagatggtg atagcgtgag cggtggc

27

<210> 23  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide sequence

<400> 23  
tcaacctttt gccccaaacaa ccactc

26

<210> 24  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide sequence

<400> 24  
tctccatcca cgctttcctg aatgtcc

27

<210> 25  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide sequence

<400> 25  
ggagaaggaa gctaaaaatc atattgtgg

29

<210> 26  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide sequence

<400> 26  
atgatgatcc taagtctacc cttttgcac

29

<210> 27  
<211> 3122  
<212> DNA  
<213> Arabidopsis thaliana

<400> 27  
ttcatatatatt cgaccttcttc ttctcattct tgcattccaaa agacacaaca agccgccato 60  
gctttccctc acaactctca ctgcaccacc gcccgtctc tcacttactc ggcttcatcg 120  
ctctcatcgcatctctcaa catactcgac ctgcgtat cactcgagct cgccgcttct 180  
caccgcctct ccatcgtcac cgccgtctcc ctctctccaa ggaaacaact cgagctctcc 240  
atttcactca ctgcacctct accaccaagg cggttccacc acttcttagct cttaaccact 300  
cgaccacacctt caccatcaac caatcaaattt gtttctctt ccattaaagc ttgacataact 360  
cgaccgctga acacttatca ctttcaagct cctcatctct tcatcggttc caacaccgct 420  
gctctcatcc cccacgaaag cttgtcatca cctctctactc atcaccagtt cactcgattc 480  
agcaacccaaa ctgcacctcg tctctcttc cactcatagt cactcgatct ctccctcacca 540  
tcttcatcatat ctcccttactt cgaccacggt gcgtctcgct ccaccattgc cattaaaaag 600  
ctcaactcgat tgtaaagag aagaagagt aagctcaacc accggcactc gaccgcgttt 660  
ccctctacac attcaacact cgaccacggt gctaccatct ccacaccgc tcttggcac 720  
catacaactcg accaacaact ctcaaaagtaa aaaaaaaaaaag aaaaaaaaaaag tcaaaaccga 780  
cagtttcaact caaccggttt actcgacccg tacgctgggt tagattgtgt ttttgggttt 840  
gctattacta acatattaac gtttatcttt gagttcgct tggttttagg tttcatcatg 900  
agtaactaca gtggaaaatc ctctatggac cctgattata atgtggatga agctaagtcc 960  
tggtccacta gaccggagtg agagcaacat gtttacgaga gctatagggaa tgaatttcaa 1020  
cgctctcgacatctcgatcttgcacgtaa tcaaagaaga gctgaaatcg cttagggaaaa gaggcgatg 1080  
tcgagtagat atgagctgat tgatgaggat atcaaaaactg agttagtgcc agagtcatgg 1140  
cgcaaggaga cgaagctact gaacaaatcc gacggggatc cagtggagga gtatatcaga 1200  
ttctttgaga tgaatgactt ctggggaaacg aggtatccct gatatgagac tttagccag 1260  
ttgggggttac tggaggacgt gcacatctg ttccgagaagt gtcatctgat aaggaggaga 1320  
caatcgagtt tctttccaca ctgcaagtgg aaatgtatga gggactcaca gactttgagc 1380  
tggataccat ggggttaggc ttcttgacgt tcttagtgg tgaacagcgg taccagattt 1440  
agatcaagaa attggaaagaa ctgtttgggtt tccctagtgg aaagggaaacc aaccccgatgt 1500  
ttgacagggaa agagcttaag gatttgggg ctactattgg gaacaatcta ccgctaaact 1560  
cgacgcgtc caagagcaac caaatccgga gtccgtgtat tcgctacttt cagcgctcg 1620  
ttgccaatgt ttttactcc agggagtcta caggcaccgt gtctaacaca gacatgaaga 1680  
tgatagattc agcgcttata gggattctcc gccttacaaa aggaaagaat gtcctgagag 1740  
gagatctaa cgactcacca ccagtaatgc ctctgttgat ccatctgtgt ggttacatga 1800  
agtgggcgtc gacaaacggc aagaagaagg taagaggagc actatgcgtg ggtggcgtt 1860  
tgacgcaat tctgaaagtt tggatgttc cgctcaagga agtagggatc gcaccgagaa 1920  
tgatggactt ggatcacttg cgccgtatgtg agttctctga gtttgacatg gttggcgtact 1980  
ttcacccgcta caggttcgag cattcatcga ttagaattcgc caacattctt ttccccctgca 2040  
tttacgctac tagattctc gagggcagga acattgactt caagcctgacg cttgaagatc 2100

tttatttcga gggcagtccg ccaactgagg agattagtca caccgaagga gctacaatag 2160  
aagatgtta tgagacatat gatatacatg aggccggagtt tgacacgagc atgtatcatt 2220  
tcagttagca tatacctcca gcgagggaaa gcaagagttt gagcgaagct cacaggaaca 2280  
acagcaagct gcagaagtgg tgcaagaaac aggataagtt actcgccaag tgccctcaggg 2340  
ctatcaagtt tctgaaggac aagatcagct gctcctcttc cactacaact attccgcaat 2400  
gacagctccc tcaggacatg cttcgagga gatatgacgc gcccggcct agagagcaga 2460  
agattctgca tgtccctgca aggcatatcatt cattcgagcc tcgtgaatct aggaagaata 2520  
ggagaacgac actcaactcga tctagcagca ggagcagacg acttctgcag tctcgttagtt 2580  
tacgcgaccc cggtgctggc cgcaatagaa gaagagaggt cgagtatctt cagagcggtg 2640  
ctggccgcca cagagctgat gagatcagttt acccacatgc tggagctgat acggaacatg 2700  
gcgggttcgtc tatggcttgg gagcaatcac aggagccat tgactaccaa ctccgttcat 2760  
tattcgactg aggttaagcgc ctcacttcac cattatatta tatcatctt tgtgatttgt 2820  
tctttatttt gtttcagtga ttggattttgt cctgagttact ctcttccaag tttattcaca 2880  
cagtggactg tgtgatttaa gtttggggga gggctcagga agtatgttgc attgtatata 2940  
tttttaagtc tgcatttcata taaggcatag aaaaacccaaa aaaaattaa aaatttcaga 3000  
aatatgttttccat acaaaaaaaag agtgttcatg tagttgcatt acattttagga tcaagtctag 3060  
agtgtttcat tttaggattgt tgcataatgca tagggataa tgatgagata gccttgcataag 3120  
ca 3122

- ET 327 548 479 US

10/088384

IC10 Rec'd PCT/PTO 15 MAR 2002

31147  
SEQUENCE LISTING

<110> Syngenta Participations AG

<120> Transcriptionally silenced Plant Genes

<130> S-31147A

<140>

<141>

<150> GB 9921964.4

<151> 1999-09-16

<160> 26

<170> PatentIn Ver. 2.1

<210> 1

<211> 2512

<212> DNA

<213> Arabidopsis thaliana

<400> 1

tcggcgttcc agcttcgcat actctctcac cgtctctcggt ttcactcgac cacttcacac 60  
ttgcgttcaa catttcgcc ggagtttctc gccattgtcc gtgcgttccgt catctccgtt 120  
caactcgaccatcccgaccggc ttaccatct ctcaactatc caccattcac tcgacacctgc 180  
cattcaactgc gcctccatcgtctttac tcgactgctc ctcaaaaccgc caccgttcc 240  
tctaattcg cggttactc gaccacactg ttacgtctct cattcgtgtta cagtcgaccg 300  
ctatacccgaa agccacaata tcactctact cgaccgttcc actcgatcgc gtacttgact 360  
ggtttagtgt gtgtgtttat ttgaactaac atattgatat ttgggtttaa gttacatct 420  
ttttcaggga atcaaatatgaa gcaactacag tggcgaatcc tccatggatg cggattacaa 480  
cgatcgatgaa gctgaatctt ggtcaacttag accagagaga gagcaacagg cttatgagag 540  
cttcagagcc gagacccaac gctcgtatgc tcgacgcaat gaaaggagag ctgagattgc 600  
tagagaaag agagcgatgaa ccagcagata tgagttgatc gacgaagata ttgacgtcga 660  
gtatgagcct gagtcatgac acagagaaaac aaaactgttg aacaagctg atgaagttac 720  
agtggaaagag tacatcagac ttttcgagct gaacgacttc tggggagcga ggtacccctg 780  
ttatgagact ctggcccagc tttaggctact ggaggacgta cagcaacttat tcgagaagtg 840  
ccatctttagt acgctgtatgt cttaccggta cgctcgatc aagaaggaaa caatagagtt 900  
tctctccact ctgcgtatgg agttgtatca gggacttact gcaatggatc tgagagatg 960  
agggttggga ttcttgactt ttccgtttaa cgagcagctt taccatgtt ctatcaagag 1020  
cttggaaagga ttatgggtt ttcccgttgg aaaggaaact aaacccaagt tcgaaaggaa 1080  
agagttaaag gatttgggtt taaccattgg gaacgatttg gcgcgtcaact ctgcaagggtc 1140  
taagagcaac cagattcgaa gcccgtgtat ccgtctactat cagcgtctcg tagcgaatgt 1200  
tctgtacccc agggaaatcta caggcaccgt gtctaaacaca gacatggaga tgattgattc 1260  
tgcactcaag ggtattctcc ggagaacaaa gggaaagaag gtcctaaagg ggcaccta 1320  
tgatcaccca cccgtcatgc ttctgttgc tccatgtgtt ggatcacgaa agtgggcgca 1380  
caccacggg aggaagaagg tgcgaggagc cttttgtgtt ggtggcgtt tgacaccgat 1440  
tctgattgca ttgtgtgtatc ctctcacgtc tccagggtt gatccgaggat tgatggattt 1500  
agatcaactt cgtcggtgtt agtttctggta gtacgacatg gttggcgatt tctatcgcta 1560  
caaattcgag cactccctgaa cccgaacagc caacattttt cttccctgca tcgaggccac 1620  
aaccataactt tagggtgaga acattgactt cagacctgctc cgtgatttacc tctactttga 1680  
gagcaactcca cccgtatgtt acaatgtccc tacgacggaa gctacagaggatgattttgc 1740  
tgagacggat gaggataggg aggaggatgt tgatcgttgc atgtatcatt tcagttagca 1800  
cgatcctcca ggcgtggaga gcaagagctt gatgtatgtt cagacggatc acagtaagtt 1860  
gcagagggtgg tgcaagaacaa aagatgtt acttcaag tgcttcaagg ccatcagtt 1920  
tctaaccggac aagataagg tcttcttcc taccacgtt attccggcagg gagagcgtcc 1980  
tcaggacatg ctttcgtatgg aatgtatgttgc gcccggccatc agtcatcaca ggcctgagcc 2040  
aagtaccac accgttgcgtt ctagtgaccg agtagtacca ccagtcctgtt caaggcattc 2100  
atcattcgatg cctcggggagc tcggggagaaa gaagaaggatc gcactcgatc ggtctggcag 2160  
caggagtaga cgtacttctcc agtccgtatc tttacgtatgc cgcgggtgtt gccgcagcag 2220  
aagaagagatc gtcgtatgtt atcagagcggtt gtcgtggccgc ggcgaaggatc gagaggtcga 2280

31147

gtaccccccag	gggaaagctg	agacacaaca	gggagattct	tcgatggcct	gggagcaatc	2340
acaggcagct	attgacgacc	aactccgctc	cttcttccac	tgaggatatgc	acctcactcc	2400
accattgtaa	tataccatct	cttgttttt	attttgtttt	tgtatgtgt	tttgtcctga	2460
gtactcttct	ccaaatttgg	tcacacagt	gactgtgtga	tttaagtttgg	gg	2512

<210> 2  
<211> 1997  
<212> DNA  
<213> *Arabidopsis thaliana*

<400> 2

ccggcagggt	caacaggctt	atgagagctt	tagagcttag	acccaacgct	cagtagctcg	60
acgcaatgaa	aggagagctg	agattgctag	aggaaagaga	gcaatgacca	gcagatatga	120
gttgatcgac	gaagatattt	acgtcgagta	tgaacctgaa	tcatggcaca	gagaaaacgaa	180
gctgttgaac	aagcccgatg	aagtacagt	agaggagtagc	atcagacttt	tcgagctgaa	240
cgacttctag	ggaacgaggt	acccctgtta	tgagacttta	gcccgagctgg	ggctacttgaa	300
ggacgtacat	cacttattcg	agaagtgc	tctggagacg	ctgatgtctt	acccgtacgt	360
cgcttacaag	aaggaaacaa	tagagttct	ctccactctg	caagtggaga	tgtatcagg	420
acttactgca	gatgagctgg	agatgtgaa	tttgggggtt	ttgactttt	cagttaacga	480
gcagcgttac	cagctatcta	tcaagagctt	ggaaggat	tttgggtttc	caagtggaaa	540
gggaactaaa	cccaagttcg	agagggaaa	gttgaagat	tttgtttaa	ccattggaa	600
cgatttggca	ctcaactctg	caaggctaa	gagcaaccag	attcgaagcc	ctgtgatccg	660
ctactatcg	cgctcagtag	cgaatgttct	gtaccccagg	gaatctacag	gcatcgtgtc	720
taacacagac	atggagatga	ttgatgttgc	actcaagggc	attctccgga	gaacaaagg	780
gaagaaggtc	ctaaagggcg	accttaatga	tacaccac	gttatgttcc	tgttgatcca	840
cctgtgtgga	tacagggat	gggcgcacac	caacgagaag	aagaagggtc	gaggagcc	900
tttgttaggt	ggcgttgtga	caccgattct	gattgcatt	gttgtaccc	tcacgtctcc	960
agggtttagt	ccgaggatga	tggattttaga	tcacttgcgt	cgttgtgagt	ttctagagta	1020
cgacatggtt	ggcgattttc	atcgctacaa	attcgagcac	tccctgaccc	gaacagccaa	1080
cattttgctt	ccctgcac	aggccacaac	cataactcag	ggtgagaaca	ttgacttcag	1140
acctgcgcgt	gattacctct	actttgagag	cgctccacc	actgatgaca	atgtccctac	1200
gacggaaagt	acagaggatg	atattgtga	gacggatgag	gatagggagg	aggagatg	1260
tacgacatg	tatcatttca	gtgacgcgt	acctccac	cgggagagca	agagcttgag	1320
tgaagctcac	agaaacaaca	gtaaatgtca	gagggtgtc	aagaaaaag	ataggctact	1380
tatcaagtgc	tccaaagcc	tcacgttct	gacggacaag	ataagctgct	tctcttctac	1440
cacagtatt	ccgcaggag	agcatcctca	ggacatgc	tcaaggagat	atgacgcgc	1500
agcgccaagt	catcacaggg	ctgagcca	tcaccacagg	cctgagcc	gtgaccg	1560
agtcccacca	gtccctgca	ggcattcatc	attcgagct	cgggagctc	ggagaaagaa	1620
gaaggctgca	ctcgctcggt	ctggcagcag	gagtagac	cttctacag	cccgtagct	1680
acgcgaccgt	ggtgctggcc	gcagcagaag	agagggatc	gagtagatc	agagcgg	1740
tggccgcac	gaaggagcag	aggtcgat	ccccacggg	aagctgag	acaacagg	1800
gattttcgat	tggcctggga	gcaatcacat	gcagctt	acgacca	ccgctcc	1860
ttccactgag	gtatgcac	cactccacca	ttgtatata	ccatctt	tttttat	1920
gtttttgtga	tgtgtttgt	cctgagtagt	ctcttccaaa	tttggcaca	cagtggact	1980
tgtgat	ttttgggg					1997

<210> 3  
<211> 1682  
<212> DNA  
<213> *Arabidopsis thaliana*

<400> 3

cttatattat	gggttggat	gtgtttaaag	aaaaggggaa	attcattgtt	gataaggaaa	60
ggaaagaat	tctagggaa	gtaagctaa	gaagttagaa	aaaatctgt	aaagggttt	120
ggaatgttaa	aaaaaagaat	gaggcttctt	ttagctttaa	aataagggtt	aaaaggctt	180
gtttttaaag	aaaaaaaaaa	aaacaggaac	cttagtttt	aaagaaatcc	aaaccgc	240
gatgtatcaa	gagcgtttag	aaagcttctc	ctagatgtt	gagaaaagaa	aagaatata	300
tgaaaaagag	tttggaaat	tcatgatgtc	aaagggtt	gtttagtgg	gacaggat	360
gtttttacca	tttagaacttc	attgttatac	tctgggtt	tggatctt	tctctgtat	420
cataatttgg	gacttac	tagcattcta	ctaaagctca	atcattctt	agggatccc	480
tgttacttaa	gccttattcta	taagggacca	tctttgtt	ttgaccc	ccttggccg	540

31147

atgagttcat	tgtatgtatgc	ttgcttgatt	cgcgttccag	aactaatgaa	tgttaaaggg	600
attggtagat	ttgaaaagcat	gtgttaggtcg	agtataagag	acggattgat	tgaaaacaag	660
gcatggctaa	cgtttttag	tagaattcaa	tcatatcgca	tcttagaact	accaacttgg	720
acattgattt	tatggctct	atcatatgct	ttggttttag	gtcccccct	tcactcct	780
ccttcaacta	tgtcttctta	tttgctttag	ggcaagcaaa	gactaagttt	gagggagttg	840
atatgtctat	aatttgcatg	ttttcagtgt	ccattcatca	tcgttttag	tccagttcg	900
tatcattcat	cactgtttta	tatcatttct	catcattctt	gcataacttgg	catgatttag	960
ataactttgc	atacatattt	catttcttag	ttgtttttag	gtgatttgg	gctgtttgca	1020
agcaaattgg	aagaatggag	ccagaaccag	aagacataact	cgaccctag	gtcgagtgac	1080
tttggggcca	tttttcccac	atactcggcc	cccaggtcg	gtgactttgg	agccatttt	1140
cccatccact	cgaccacccgg	gtcgagtaac	cttagtctcg	gccactcgat	gacactactc	1200
gaccggcagg	cgagtatcac	ttcgccacac	cacctgacaa	cactcgacca	atcaacttac	1260
caagttactc	gacccctgg	tcgagatatca	tcactcacca	ccatcagcat	cactcgaccg	1320
gacactcgat	cacgtctca	cagtctactc	aaatccgcag	tcaaccagac	aagctgagca	1380
caaggaagag	aagaggagaa	gacaaggatgc	ttggaaaggcg	cctggacctc	catcgatca	1440
cgaagccat	ctcgccccat	tatctctcta	ttggccgagc	gattaggtt	ttggccggtc	1500
tactatcatt	ttatttcggt	ttgtataaaat	agatgtctt	gggttttgtc	ctgagacatc	1560
tagtcacat	tgagttttt	ttgcttcagt	tttattttct	gttctactct	gctgcgcgc	1620
tttgcctct	gcaacctgta	attcgagatt	tttccaagg	attcagattc	cgcatttgat	1680
tt						1682

&lt;210&gt; 4

&lt;211&gt; 1652

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis thaliana

&lt;400&gt; 4

cttatatcat	gggtttggat	cagttaaaaa	aaaaaaaaagg	gtgaattcat	tgttgataag	60
gaaaggaaaa	gaattctagg	ggaagtaagc	taaagaagg	agaaaaaaaa	aatctagta	120
aagggtttgg	gaatgttaaa	gaaaagaatg	aggttcttg	tagctaaaga	agaagggtt	180
aaagcctttt	gttttaaaga	ttaaaaacag	gaaccttagt	tgttaaagaa	atccaaatac	240
gctagatgt	tcagagtgtt	gagaaggctt	ctcctagat	taagagaaaa	gaaaagaatg	300
atatggaaaa	gagtttgaaa	gattcatgag	tgc当地gggt	agatgttaat	tcttgtattt	360
ggacttgggt	tgggattacc	attagagctt	cattgttata	ctatggtag	atgggatttt	420
atctctgtat	gcataactt	ggacttaccc	ttagcattt	actaaagctc	aatcatttt	480
gagagatccc	ctgttactt	agcctattt	gtaagggacc	atctttgtct	cttgaccc	540
accttagcca	aatgagttc	ttgatgtatc	attgtttt	tcacgttcca	gaactaatga	600
atgttaaagg	gattggtaga	tttgaagca	tgttaggtc	gagttataaga	gacggatgt	660
ttgataacaa	ggcatggct	acgttttgc	gtaaaatttca	atcatatcgc	atcttagaac	720
taccaactt	gacattgtt	ttatttgct	tatcgatgc	tttggttctg	agtccccacc	780
ttcaaaacctc	tcccttcaact	atgtcttctt	atttgctt	ggcaagcaa	agactaagtt	840
tgggggagtt	gatatgtcta	taatttgcat	gttttgcgt	tccatttc	atcggtttga	900
gtccagttt	gtatcattca	tcactgtttt	atatcattt	tcatcattt	tgcatactt	960
gcatgattag	gatagcttt	tatacatatt	gcatttctg	gttggtttta	gttgatttgg	1020
agctgtttgc	gagcaattt	gaagaaaacga	gccagaacaa	gaaggccatac	tcgacccccc	1080
ggtcgagtga	cttggagcc	attcttccca	tctactcgac	ccgggggtcg	agtaaccc	1140
gctcaggcca	ctcgatgacg	ccactcg	ccccctggcg	agatactt	cgccacacca	1200
cctgaccaca	ctcgcccggt	cacttacca	cgttactcg	ccccctggtc	gagttatcatc	1260
actcaccacc	aacaccatca	ctcgaccggg	cactcgatca	catcttata	gtctactcaa	1320
atccgcactc	aaccagacaa	gctgagcaca	aggaagagaa	gaggagaaga	caaagtgtt	1380
ggaaggccgg	ttgacctcca	tcggatcacg	agaagccca	tctcgccca	ttatcattt	1440
atggggccgg	cgatttaggtt	attggccctgt	ctactatcat	tttattttgt	tatgtataaa	1500
tagatgtctt	agggttctgt	accaggacat	ctagtcgaca	ttgagtttt	ttgcttcagt	1560
tttattttct	gttttctctg	ctgcgcgcgt	tttggttctg	caacctgtaa	ttcgagattt	1620
ttccaagtt	ctcagatttcc	gcatttgatt	tc			1652

&lt;210&gt; 5

&lt;211&gt; 903

&lt;212&gt; DNA

&lt;213&gt; Arabidopsis thaliana

31147

<210> 6

<211> 614

<212> DNA

<213> *Arabidopsis thaliana*

<400> 6

gaggggcaagc	aaagactaag	tttgggggag	ttgataagtg	tgtatttg	atgtttgag	60
catccatttg	tcatcacttt	agcatcatat	catcaactgtt	ttataccatt	tcacatcatt	120
tgtcatca	ttgcatgttt	aggatagttt	tgcatacgatg	ttgcatat	gcgttgattt	180
cagggtattt	ggagctgttg	acgagctatc	tggaaagaagc	agacctgatc	atgacaacc	240
actcgacc	gaggtcgagt	aggagcttca	agatctcaag	agactactcg	acaaccagg	300
cgagtagagc	acatcaccac	ttcacccatc	cactcgacc	cgagggtcgag	tgcacatc	360
tccatcacca	gacggtact	cgatcacttc	actcgacctt	gaggtcgagt	gtcttcacct	420
ccatcatcag	acaaccactc	gacctctca	ctccacctag	aggtcgagta	tctccatctt	480
accactcgac	tgcataactcg	atgacaagct	tcagagcctt	cttaattccg	cactcaacca	540
gacactcgag	cacaaggaag	aaaagaagac	tccagctatt	cactcgagct	ctcactcgac	600
cacgtgggtc	gagt					614

<210> 7

<211> 1956

<212> DNA

<213> *Arabidopsis thaliana*

<400> 7

ttttgttca	ccggattAAC	tcatgtct	cgttgcgt	tgtgtttGC	gtatgtaaTG	60
aatttgaag	aaaactgaac	catgcctaga	ttgctctact	cgaccacact	gtcatgatCT	120
gataccattc	cctatcaatt	tgaacctgaa	tttgatcttt	aattatcatG	tctgcataaA	180
atttgaactc	atggataccC	taaaaacttC	ggatTTCTT	attcattttG	atcaCTTGT	240
ttaatccaag	tagtgcACt	tccctttag	agcaggtaAC	ccgaacccAA	acttagactCT	300
ttttcaagc	cttatacAC	tcgtgaggGT	tttgaggtc	tatTCCTGAT	tcaGCTTGGT	360
agaaaagtTT	aggTCgtAA	cgacagAGAT	agtgnctCAT	gtatTCCTAG	tTCGCACTTT	420
ttggactaga	taggactGGG	tgggcgttA	tactttAGGT	tgggatGNGT	ttaaaaagAAA	480
aaaaaaaggGGG	ttgtattcATT	gatgagaaaa	ggtaaaAGAC	tctaggtGA	gtaagataAA	540
gaagcagaaaa	aggTCTAGTA	aaggTTTGG	gatttGtaAA	aaaaagaaAG	agttCTTGT	600
agcttattgaa	gatgggcaAA	agcccTcGGT	tttaaaatGT	aaaaaacAGG	aacCTTAGTT	660
gttaaagaaaa	tccaaatCCG	ctagatgtAT	caaagtGTTG	agaaaAGCTT	tcctAGAGTT	720
aagagaaaaAG	aaaAGAATGA	ttagaaaaAG	ggcttaaAGG	attcatGAAT	gcaaaAGGGTA	780
gaggtaagtt	cttataCTGG	gattggAGAT	gggatttACCA	ttagagCTTC	atctgtatATA	840
ctcttaggtAG	atgggatCTT	atctctGcat	gcataGTTG	ggacttACCT	ttagcattCT	900
actaaagCTT	aatcattttt	tgagagatCC	cctgttACTG	aagccttATTC	tgtaaAGGGAC	960
catcttGTC	tcttgacCTT	ttaccttAGC	caaAtGAGTT	cattgtatGAT	gcattGCTTG	1020
attcacgttc	cagaactaat	gaatgttaAA	gggatttGgtA	gatttGAAAAA	catgtgttagG	1080
tcgagcatat	gagtccggatt	gattgtatGT	aaggcatGGC	taaagtTTTT	cagttagaATT	1140

31147

cgatcatatac	gcagcttaga	actatcaact	tggacattga	tttcatttgg	tttatctagt	1200
gctttggctc	ttagtccccg	atttcaaacc	tcacctctag	cttgcatttta	attgtttgct	1260
tgaggcagaag	caaagactaa	gttggggga	gttgataagt	gtgtattttg	catgttttga	1320
gcatccattt	gtcatcaact	tagcaccata	tcatcaact	tttataccat	ttcicatcat	1380
ttgtcatcac	tttgcatgtt	taggatagtt	ttgcattgt	gtggcatatt	tgtgttgtt	1440
tcaagtgtt	cggagctgtt	gaagaactaa	tttggagaag	cgacactgtat	catgccaac	1500
cactcgacct	caggctcgat	agacgcttca	cgacctcaac	acaccactcg	accacactgt	1560
cgagtgtagg	acttcaccac	ttcacctcat	cactcgaccc	cctggccgag	taccccacga	1620
gagtcaactg	atcacttcac	tcgaccccca	gtcgagtt	cttcacctcc	accacactgac	1680
catcaactcg	tcacacgact	ctacctggaa	gtcgagttatc	accatcacca	ccactcgact	1740
acataacttga	tgtcgagctt	cagacttcc	tccattccgc	actcaaccag	acactcgagc	1800
acaaggaaaa	aaagaagat	ctagcttac	actcgacc	tcactcgacc	acctgggtcg	1860
agtagacttc	ttaatccgtc	tcaatactgc	gtcgtttga	gtatttaggt	ttcggaatat	1920
ttttgtata	agtagcacgt	actttacatt	ttcgag			1956

<210> 8

<211> 2105

<212> DNA

<213> Arabidopsis thaliana

<400> 8

ttagcatttt	gttcaactag	ataaaactcag	tgccctcggt	gttagttgtc	tgatgcatac	60
tcaatgaaat	tgaagtaaaa	ctgcaccatg	cctagattgc	tctactcgac	cacactgtta	120
ggatctgata	ccattcccta	tcaatttga	cttgaatctg	attttagaatt	atcatgtctt	180
gccatcgaaat	ttgaactcat	ggatacccta	aaatacttgg	attttcttac	tcattttaac	240
cactcttgtt	atccaaatg	gctgactctc	cttatttagag	cagttaccc	gaatccaaac	300
ctaatttttc	tttcgagccc	tatattca	gtgagttt	gtgaggtctt	atttcaattt	360
agcttggtag	aaagtgttag	gttcgtaa	acagagatag	tgtctcatgt	agttctagtt	420
cgcgtttttt	ggactggata	ggacttaggt	ggcgcttata	tcatgggtt	ggatcagtt	480
aaaaaaaaaa	aagggtgaat	tcattttga	taaggaaagg	gaaagaattt	taggggaagt	540
aagctaaaga	agtttagaaaaa	aaaaaaatct	agtaaagggtt	ttggaaatgt	taaagaaaaag	600
aatgagggtt	tttttagctt	aagaagaagg	gttaaaagcc	tttttttttta	aaagattaaaa	660
acaggaaacct	tagttgtttaa	agaaatccaa	atacgctaga	tgtatcagag	tgttgagaaa	720
gcttcctcta	gagttaaag	aaaagaaaat	aatgtatgat	aaaagagttt	gaaagattca	780
tgagtgcataa	gggttaggtt	aaattttgtt	attggactt	gagttggat	taccatttta	840
gcttcattgt	tatactatgg	gttagatggg	ttttatctt	gtatgcataa	tttgggactt	900
accttttagca	ttctactaaa	gctcaatcat	tctttagaga	tcccctgtt	cttaagcata	960
ttctgtttaa	gaccatctt	gtcttgc	ttcacctttag	ccaaatgagt	tcattgtatg	1020
tgcattgtt	gattcacgtt	ccagaactaa	tgaatgtttaa	agggtttt	agatttggaaa	1080
gcatgtgtat	gtcgagttat	agagacggat	tgattgtttaa	caaggcatgg	ctaactttt	1140
cgagtaaaat	tcaatcatat	cgcatctttag	aacttccaa	ttggacattt	attttttttt	1200
ctctatcaga	tgctttgggtt	ctgatcccc	accttcaaa	ctctcccttca	actatgtctt	1260
cttatttgc	tgaggggcaag	caaagactaa	gtttggggga	gttgcattgt	ctataattt	1320
catgtttca	gtgtccatcc	atcatcgat	tgagtccat	ttcgtatcat	tcatcaactgt	1380
tttataatcat	ttctcatcat	tcttgcatac	tttgcattat	taggtatgt	ttgtatacat	1440
attgcatttc	tgagttgtt	tttaggtatt	tggagctgtt	tgcgagcaaa	tttggaaagaaa	1500
cgagccagaa	caagaagcca	tactcgaccc	cctggtcgag	tgacttttgg	gccatttttc	1560
ccatctactc	gaccgggggg	tcgatcaacc	tcaactcgat	ccactcgat	acgcccactcg	1620
ccccccctgg	tcgatataca	cttcggccaca	ccacatcgacc	acactcgcc	gttcactcta	1680
ccacgttact	cgacccttgc	gtcgatgtt	atcaacttcc	accaacacca	tcactcgacc	1740
gggcactcga	tcacatcttc	atagtctact	caaattccgca	ctcaaccaga	caagctgagc	1800
acaaggaaga	gaagaggaga	agacaaatgt	tttggaaagcg	gcctggaccc	ccatcgatc	1860
acgaagaagc	ccatctcgcc	ccattatcat	tctatggggcc	ggggcattag	tttattggcc	1920
cgtctactat	cattttat	cgatgtat	aaatagatgt	cttagggttt	tgttaccagga	1980
catctatcg	acatttgat	tttttgc	agttttat	tctgttttct	ctgctgcgccc	2040
gttttgc	ctgcaacctg	taattcgaga	tttttccaag	ttattcagat	tccgcattt	2100
atttc						2105

<210> 9

<211> 4860

<212> DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: virtual TSI

&lt;400&gt; 9

tcggcgttcc	agcttcgcac	actctctcac	cgtctctcg	ttca	tcgtac	cacttcacac	60
ttcgcccaa	catcttcgccc	ggagtttctc	gccattgtcc	gtgcttccgt	catctccgtt	120	
cactcgacca	ccggaccggc	ttcaccatct	ctcaactatc	caccattcac	tcgacctcgc	180	
cattcactgc	gcctccattc	gtctcttac	tcgactgctc	ctcaaaccgc	caccgtcttc	240	
tctaaattcg	ccgtttactc	gaccacactg	ttacgtctc	cattctgtta	cagtgcacg	300	
ctataccga	agccacaata	tcactctact	cgaccgtttc	actcgatcgc	gtacttgact	360	
ggtttagtgt	gtgtgtttat	ttgaactaac	atattgtat	ttgggtttga	gttacattct	420	
ttttcaggga	atcaaatatga	gcaactacag	ttggcaatcc	tccatggatg	cggttaccaa	480	
cgtcgatgaa	gctgaatctt	ggtcaactag	accagagaga	gagcaacagg	cttatgagag	540	
cttcagagcc	gagacccaac	gctcagtagc	tcgacgcaat	gaaaggagag	ctgagattgc	600	
tagagaaaag	agagcgatga	ccagcagata	tgagttgatc	gacgaagata	ttgacgtcga	660	
gtatgagcct	gagtcatggc	acagagaaaac	aaaactgttg	aacaaggctg	atgaagttac	720	
agtggaaagag	tacatcagac	ttttcgagct	gaacgacttc	tggggagcga	ggtacccctg	780	
ttatggagact	ctagcccgac	tttagctact	ggaggagcgt	cagcacttat	tcgagaagtg	840	
ccatctttag	acgctgtatg	cttaccggta	cgtcgcttac	aagaaggaaa	caatagagg	900	
tctcttccact	ctgcaagtgg	agttgtatca	gggacttact	gcagatgaac	tggagagtg	960	
agggttggga	ttcttgactt	tttcagttaa	cgagcagcgt	taccagctat	ctatcaagag	1020	
cttggaaagga	ttattinggtt	ttcccagtgg	aaagggaact	aaacccaagt	tcgaaaggga	1080	
agagttaag	gatttgggtt	taaccattgg	gaacgatttg	gcgcctaact	ctgcaagg	1140	
taagagcaac	cagattcgaa	gcccctgtat	cgcgtactat	cagcgtctag	tagcaatgt	1200	
tctgttcccc	agggaaatcc	caggcaccgt	gtctaaccata	gacatggaga	tgattgattc	1260	
tgcactcaag	ggatttctcc	ggagaacaaa	ggggaaaga	gtccttaagg	gacacccctaa	1320	
tgataccca	ccggcatgtc	ttctgttgc	ccatatgtgt	ggatacagga	agtgggcgca	1380	
caccaacggg	aggaagaagg	tgcgaggagc	cctttgtgt	ggtggcgtt	tgacaccgat	1440	
tctgatttgc	tgtgtgttac	ctctcacgtc	tccagggttt	gatccgagga	tgtatggatt	1500	
agatcacttg	cgtcgttgc	agtttcttgc	gtacgacatg	gttggcgtt	tctatcgct	1560	
caaattcgag	cactccctga	cccgaacagc	caacattttg	cttccttgca	tcgaggccac	1620	
aaccatctt	tagggtgaga	acattgtact	cagacgtcgc	cgtgattacc	tctactttg	1680	
gagcactcca	ccgactgtat	acaatgtccc	tgacgcaaa	gctacagagg	atgatttgc	1740	
tgagacggat	gaggatagg	aggaggagta	tgatacgtac	atgtatcatt	tcagtgagca	1800	
cgtacctcca	gcgcaggaga	gcaaggagctt	gagtgaagct	cacagaaaca	acagtaagtt	1860	
gcagagggtgg	tgcaagaaac	aaagataggtt	acttatcaag	tgcttcaagg	ccatcacgtt	1920	
tctaaccggac	aagataagtt	gcttcttcc	taccacagt	attccgcagg	gagagcgtcc	1980	
tcaggacatg	ccttcgaaga	gatatgacgc	gccaggccca	agtcatcaca	ggcctgagcc	2040	
aagtccaccac	aggcctgac	ctagtgcac	agtagtacca	ccagtcctg	caaggcattc	2100	
atcatcgag	cctcggggagc	tcggggagaa	aaagaaggct	gcactcgct	ggctggcag	2160	
caggagtaga	cgacttctcc	agtcgggtat	cttacgcgtc	cgcgtgt	gccgcagcag	2220	
aagaagagag	gtcgagtatc	atcagagcgg	tgctggccgc	ggcgaaggag	cagaggtcga	2280	
gtaccccccag	ggggaaagctg	agacacaaca	gggagattct	tcgatggcct	gggagcaatc	2340	
acaggcagct	attgacgacc	aactccgctc	tttcttccac	tgaggtatgc	acctca	2400	
accattgtaa	tataccatct	tttgtttttt	attttgttt	tgtgtatgt	tttgtccgt	2460	
gtactcttct	ccaaattttg	tcacacagt	gactgtgt	ttaagtttt	ggggagg	2520	
caggaagtgt	gtgtgtcatt	gtataatctt	gagtttgcatt	tcatctaagg	catagaaaa	2580	
ccaaaaaaat	tgaaaaaaatc	cagaaaaatga	tttcacaaa	atagatgtt	catgtatgt	2640	
cattgcattt	aggatcgagt	ttagagtgtt	tcgtttagga	tttgtgcata	tgcataggg	2700	
ataataatga	gatagccctg	taagcatttt	ggttccacag	ataagctcag	tgccctcg	2760	
gttagttgtt	tgatgcgtt	tcattgtaaat	tgaagtaaga	actgcacgt	gcctagattt	2820	
ctctactcga	ccacactgtt	aggatctgtat	atcattccct	atcaatttga	acttgaatct	2880	
gatttagaat	tatcatgtct	tggcatcgaa	tttgaactca	tggataccct	aaaatactt	2940	
gattttctta	ctcattttaa	ccactcttgc	tgatccaat	agctgactct	ccttattaga	3000	
gcagttaaacc	catacccaa	cctgaacttt	ctttcaagcc	ctatatact	tgtgatgtt	3060	
tgtgagggtct	tattingt	gagcttggta	gaaagtgtt	ggttgcgtac	gacagagata	3120	
gtgtctcatg	tagttctat	ttgcgttctt	cagactggat	aggacttaggt	gggcgttat	3180	
atcatgggtt	gggatgtgtt	taaaagaaaa	gagggatct	attgttgcatt	agggaaagg	3240	
aagaattcca	ggggaaagtaa	gctaaagaag	tttagaaaaaa	aatcttagaa	aggttttgg	3300	
aatgttaaag	aaaagaatga	ggttcttgc	agctaaagaa	gaagggtttaa	aaggccttgg	3360	
ttttaaagat	aaaaaaaaaa	acaggaacct	tagttgttaa	agaaatccaa	acccgcgtaga	3420	

31147

tgtatcaaga	gcgttgagaa	agcttccct	agagttaaaa	aaaagaaaa	gaatgatatg	3480
aaaaagagtt	tgaaaagattc	atgagtgc当地	aggtagatg	taagttggga	caggagttgg	3540
ttttaccatt	agaacttcat	tgttatactc	tgggttagatg	ggatcttac当地	tctgtatgca	3600
taatttggga	cttaccttta	gcattctact	aaagctcaat	cattcttgag	ggatccccctg	3660
ttacttaagc	ctattctata	agggacatc	tttgtctctt	gacccctacc	ttggccgaat	3720
gagttcattt	atgatgcatt	gcttgattcg	cgttccagaa	ctaataatg	ttaaaggatg	3780
tggtagatg	gaaagctatgt	gtaggtcgag	tataagagac	ggattgttgg	aaaacaaggc	3840
atggctaa	tttttgagta	gaattcaatc	atatcgcatc	ttagaactac	caacttggac	3900
attgatttta	tttgctctat	catatgcttt	ggttttgagt	ccccgccttc	actcctctcc	3960
ttcaactatg	tcttcttatt	tgcttgaggg	caagcaaaga	ctaagttga	gggagttgt	4020
atgtctataa	tttgcatgtt	ttcagtgltcc	attcatcatc	gttttgagtc	cagtttgcgt	4080
tcattcatca	ctgttttata	tcattttctca	tcattttctgc	atactttgca	tgtttaggt	4140
aactttgcatt	acatatttgc	tttctgagg	gttttcaggt	gattttgagc	tgtttgcga	4200
caaatttggaa	gaaatgagcc	agaaccagaa	gacatactcg	acccccttaggt	cgagtgtactt	4260
tggggccatt	cttcccacat	actcggcccc	cagggtcgagt	gactttggag	ccatttcttc	4320
catccactcg	accaccgggt	cgagtaacct	tagctcaggc	cactcgatga	cactactcga	4380
ccccaggtcg	agtatcactt	cgccacacca	cctgacaaca	ctcgaccaat	cactctacca	4440
agttactcgat	ccccctggtc	gagtatcatc	actcaccac	atcagcatca	ctcgacccgga	4500
cactcgatca	cgttttcaca	gtctactcaa	atccgcagtc	aaccagacaa	gctgagcaca	4560
aggaagagaa	gaggagaaga	caaagtgc当地	ggaaaggccg	tggacccctca	tccgatcacg	4620
aagcccatct	cggcccattt	tctctctatg	ggccggcggca	tttaggttatt	ggcccgtcta	4680
ctatcatttt	atttcgtttt	gtataaaatag	atgtcttagg	gttttgtctt	gagacatcta	4740
gtcgacattt	agtttttttt	gcttcagttt	tattttctgt	tctactctgc	tgcggccgtt	4800
ttqcttcgtc	aacctgttaat	tcgagatttt	tccaagttat	tcagattccg	catttgattt	4860

<210> 10  
<211> 648  
<212> PRT  
<213> Ara

<400> 10  
 Met Ser Asn Tyr Ser Gly Glu Ser Ser Met Asp Ala Asp Tyr Asn Val  
 1 5 10 15  
 Asp Glu Ala Glu Ser Trp Ser Thr Arg Pro Glu Arg Glu Gln Gln Ala  
 20 25 30  
 Tyr Glu Ser Phe Arg Ala Glu Thr Gln Arg Ser Val Ala Arg Arg Asn  
 35 40 45  
 Glu Arg Arg Ala Glu Ile Ala Arg Gly Lys Arg Ala Met Thr Ser Arg  
 50 55 60  
 Tyr Glu Leu Ile Asp Glu Asp Ile Asp Val Glu Tyr Glu Pro Glu Ser  
 65 70 75 80  
 Trp His Arg Glu Thr Lys Leu Leu Asn Lys Pro Asp Glu Val Thr Val  
 85 90 95  
 Glu Glu Tyr Ile Arg Leu Phe Glu Leu Asn Asp Phe Trp Gly Ala Arg  
 100 105 110  
 Tyr Pro Cys Tyr Glu Thr Leu Ala Gln Leu Arg Leu Leu Glu Asp Val  
 115 120 125  
 Gln His Leu Phe Glu Lys Cys His Leu Glu Thr Leu Met Ser Tyr Pro  
 130 135 140  
 Tyr Val Ala Tyr Lys Lys Glu Thr Ile Glu Phe Leu Ser Thr Leu Gln  
 145 150 155 160  
 Val Glu Leu Tyr Gln Gly Leu Thr Ala Asp Glu Leu Glu Ser Glu Gly

31147

165	170	175	
Leu Gly Phe Leu Thr Phe Ser Val Asn Glu Gln Arg Tyr Gln Leu Ser			
180	185	190	
Ile Lys Ser Leu Glu Gly Leu Phe Gly Phe Pro Ser Gly Lys Gly Thr			
195	200	205	
Lys Pro Lys Phe Glu Arg Glu Glu Leu Lys Asp Leu Trp Leu Thr Ile			
210	215	220	
Gly Asn Asp Leu Ala Leu Asn Ser Ala Arg Ser Lys Ser Asn Gln Ile			
225	230	235	240
Arg Ser Pro Val Ile Arg Tyr Tyr Gln Arg Ser Val Ala Asn Val Leu			
245	250	255	
Tyr Pro Arg Glu Ser Thr Gly Thr Val Ser Asn Thr Asp Met Glu Met			
260	265	270	
Ile Asp Ser Ala Leu Lys Gly Ile Leu Arg Arg Thr Lys Gly Lys Lys			
275	280	285	
Val Leu Lys Gly Asp Leu Asn Asp Thr Pro Pro Val Met Leu Leu Leu			
290	295	300	
Ile His Met Cys Gly Tyr Arg Lys Trp Ala His Thr Asn Gly Arg Lys			
305	310	315	320
Lys Val Arg Gly Ala Leu Cys Val Gly Gly Val Val Thr Pro Ile Leu			
325	330	335	
Ile Ala Cys Gly Val Pro Leu Thr Ser Pro Gly Phe Asp Pro Arg Met			
340	345	350	
Met Asp Leu Asp His Leu Arg Arg Cys Glu Phe Leu Glu Tyr Asp Met			
355	360	365	
Val Gly Asp Phe Tyr Arg Tyr Lys Phe Glu His Ser Leu Thr Arg Thr			
370	375	380	
Ala Asn Ile Leu Leu Pro Cys Ile Glu Ala Thr Thr Ile Leu Xaa Gly			
385	390	395	400
Glu Asn Ile Asp Phe Arg Pro Ala Arg Asp Tyr Leu Tyr Phe Glu Ser			
405	410	415	
Thr Pro Pro Thr Asp Asp Asn Val Pro Thr Thr Glu Ala Thr Glu Asp			
420	425	430	
Asp Phe Ala Glu Thr Asp Glu Asp Arg Glu Glu Glu Tyr Asp Thr Ser			
435	440	445	
Met Tyr His Phe Ser Glu His Val Pro Pro Ala Gln Glu Ser Lys Ser			
450	455	460	
Leu Ser Glu Ala His Arg Asn Asn Ser Lys Leu Gln Arg Trp Cys Lys			
465	470	475	480
Lys Gln Asp Arg Leu Leu Ile Lys Cys Phe Lys Ala Ile Thr Phe Leu			
485	490	495	
Thr Asp Lys Ile Ser Cys Phe Ser Ser Thr Thr Ala Ile Pro Gln Gly			

31147

500	505	510
Glu Arg Pro Gln Asp Met Pro Ser Lys Arg Tyr Asp Ala Pro Gly Pro		
515	520	525
Ser His His Arg Pro Glu Pro Ser His His Arg Pro Glu Pro Ser Asp		
530	535	540
Arg Val Val Pro Pro Val Pro Ala Arg His Ser Ser Phe Glu Pro Arg		
545	550	560
Glu Leu Gly Arg Lys Lys Ala Ala Leu Ala Arg Ser Gly Ser Arg		
565	570	575
Ser Arg Arg Leu Leu Gln Ser Arg Ser Leu Arg Asp Arg Gly Ala Gly		
580	585	590
Arg Ser Arg Arg Arg Glu Val Glu Tyr His Gln Ser Gly Ala Gly Arg		
595	600	605
Gly Glu Gly Ala Glu Val Glu Tyr Pro Gln Gly Glu Ala Glu Thr Gln		
610	615	620
Gln Gln Asp Ser Ser Met Ala Trp Glu Gln Ser Gln Ala Ala Ile Asp		
625	630	635
Asp Gln Leu Arg Ser Phe Phe His		
645		

<210> 11  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 11  
tggttaccca gataagctca gtgccctc 28

<210> 12  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 12  
cttcagactg gataggacta ggtgggcg 28

<210> 13  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>

31147

<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 13  
cgcccaccta gtcctatcca gtctgaag 28

<210> 14  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 14  
cgcataacaac aactaacaac gagggcac 28

<210> 15  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 15  
cgataaacatc gaccgtattt ctcgcc 26

<210> 16  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 16  
aactagctcc catccgtctt cgacatcc 28

<210> 17  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 17  
tgcatcacac cggattggat tgac 24

<210> 18  
<211> 28  
<212> DNA  
<213> Artificial Sequence

31147

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 18  
tgttccccctg aaccatagca atgagacc 28

<210> 19  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 19  
caaacagaca gagtgtggcc caccacc 27

<210> 20  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 20  
caaacagaca gagtgtggcc caccacc 27

<210> 21  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 21  
tgcaaaccctt caggaccaag tctaccc 27

<210> 22  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
Oligonucleotide sequence

<400> 22  
acagatggtg atagcgtgag cggtggc 27

<210> 23  
<211> 26

31147

<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide sequence

<400> 23  
tcaacccttt gccccaaacaa ccactc

26

<210> 24  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide sequence

<400> 24  
tctccatcca cgctttcctg aatgtcc

27

<210> 25  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide sequence

<400> 25  
ggagaaggaa gctgaaaatc atattgtgg

29

<210> 26  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide sequence

<400> 26  
atgatgatcc taagtctacc cttttgcac

29

<210> 27  
<211> 3122  
<212> DNA  
<213> Arabidopsis thaliana

<400> 27  
ttcatatatatt cgacctcttc ttctcattct tgcattccaaa agacacaaca agccgccatc 60  
gctttccctc acaactctca ctgcgaccacc gcccgctctc tcacttactc ggcttcatcg 120  
ctctcatcgc catctctcaa catactcgac ctcgcgatat cactcgagct cgccgcttct 180  
caccgcctct ccatcggtcac cgccgtctcc ctctctccaa ggaacaact cgagctctcc 240  
atttcactca ctcgacacct accaccaagc cggcttcacc acttcttagct cttAACCT 300  
cgaccacacctt caccatcaac caatcaaatac gttttctctt ccattaaagc ttgacataact 360  
cgaccgctga acacttatca cttcaagct cctcatctt tcattcggttc caacaccgct 420

31147

gctctcatcc	cccacgaaag	cttgtcatca	cctctcactc	atcaccagtt	cactcgattc	480
agcaacccaaa	ctcgacctcg	tctctttgc	cactcatagt	cactcgatct	ctcctcacca	540
tcttcatcat	ctcccttact	cgaccaccgt	gcgtctcgct	ccaccattgc	catttaaaag	600
ctcaactcgat	tgtcaaagag	aagaagagt	aagctcaacc	accggccatc	gaccgcgtt	660
ccctctacac	attcaacact	cgaccacgg	gctaccatct	ccacacccgc	tcttgttcac	720
catacaactcg	accaacaact	ctcaaagtaa	aaaaaaaaaaag	aaaaaaaaaaag	tcaaaaccga	780
cagtttact	caaccggttt	actcgaccgg	tacgctgggt	tagattgtgt	tttgggtttt	840
gctattacta	acatattaac	gtttatcttt	gagtttgc	tgttttttagg	tttcatcatg	900
agtaactaca	gtggaaaatc	ctctatggac	cctgattata	atgtggatga	agctaagtcc	960
tggtcacta	gaccggagtg	agagcaacat	gtttacgaga	gctataggga	tgaatttgaa	1020
cgctctgcag	ctcgacgtaa	tcaaagaaga	gctgaaatcg	ctagaggaaa	gagggcgtatg	1080
tcgagtagat	atgagctgt	tgatgaggat	atcaaaactg	agtatgagcc	agagtcatgg	1140
cgcggaga	cgaagctact	gaacaaatcc	gacgagggt	cagttggagga	gtatatcaga	1200
ttcttgaga	tgaatgactt	ctggggaaacg	aggtatccct	gatatgagac	tttagcccg	1260
ttggggttac	tggaggacgt	gcagcatctg	ttcgagaagt	gtcatctgtat	aaggaggaga	1320
caatcgagtt	tctttccaca	ctgcaagtgg	aatgttatga	gggactcaca	gactttgagc	1380
tggataccat	ggggtaggc	ttcttgacgt	tcttagtgg	tgaacagcgg	taccagattt	1440
agatcaagaa	attggaagaa	ctgtttgggt	tccctagttg	aaagggaacc	aaccccaggt	1500
ttgacaggga	agagcttaag	gattttgggg	ctactattgg	gaacaatcta	ccgctaaact	1560
cgacgcgtc	caagagcaac	caaatccgg	gtcctgtat	tcgctacttt	cagcgctcgg	1620
ttgccaatgt	ttttactcc	aggaggatcta	caggcaccgt	gtctaacaca	gacatgaaga	1680
tgatagattc	agcgcttata	gggattctcc	gccttacaaa	aggaaaagaat	gtcctgagag	1740
gagatcttaa	cgactcacca	ccagtaatgc	ctctgttgat	ccatctgtgt	gggtacatga	1800
agtggcgcgt	gacaaacggc	aagaagaagg	taagaggagc	actatgcgtg	ggtggcgtt	1860
tgacgccaat	tctgaaagtt	tgtggagttc	cgctcaagga	agttagggtt	gcaccgagaa	1920
tgatggactt	gatcacttgc	cgccgatgt	agttctctgt	gttggacatg	gttggcgtact	1980
ttcacccgcta	caggttcgat	cattcatcg	ttagaatcgc	caacattctt	ttccccctgca	2040
tttacgtac	taggatttctc	gagggcaggg	acatttgacct	caaggctgcg	tttgaagatc	2100
tttatttcga	gggcagtcgg	ccaaactcgagg	agatttagtca	cacccaaggaa	gctacaatag	2160
aagatgttga	ttagacatata	gatatacgatg	aggcggagtt	tgacacgagc	atgtatcatt	2220
tcagttagca	tataccctcca	gcgagaaaaa	gcaagagttt	gagcgaagct	cacaggaaca	2280
acagcaagct	gcagaagtgg	tgcaagaaaac	aggataagg	actcgccaag	tgccctcaggg	2340
ctatcaagtt	tctgaaaggac	aagatcgact	gtcctcttc	cactacaact	attccgcata	2400
gacagctccc	tcaggacatg	ccttcgagga	gatatgacgc	gcccggcc	agagagcaga	2460
agatttcgca	tgtccctg	aggcattcat	cattcgaggc	tcgtaatct	aggaagaata	2520
ggagaacgac	actcaactcg	tcttagcagca	ggagcagacg	acttctgcag	tctcgttagtt	2580
tacgcgaccg	cgggtctggc	cgcaatagaa	gaagagaggt	cgagttatcct	cagagcgttg	2640
ctggccgcca	cagagctgtat	gagatcgagt	acccacatgc	tggagctgtat	acggaacatg	2700
gcgggtcgatc	tatggcttgg	gagcaatcac	aggcagccat	tgacttaccaa	tttcgttcat	2760
tattcgactg	aggttaagcgc	ctcaacttcc	cattatatta	tatcatctct	tgtgatttgc	2820
tctttat	tttgcgt	ttggattttgt	cctcgatct	ctcttccaaag	tttatttcaca	2880
cagtggactg	tgtgattttaa	gtttgggggg	gggctcagga	agtatgttgc	attgtatata	2940
tttttaagtc	tgcatttcata	taaggcatag	aaaaacccaa	aaaaaattaa	aaatttcaga	3000
aaatgatttc	acaaaaaaag	agtgttcatg	tagttgcatt	acatttagga	tcaagtctag	3060
agtgttcat	ttaggattgt	tgcataatgc	tagggataaa	tgatgagata	gccttgcata	3120
ca						3122